

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:06:21 ; Search time 103.196 seconds
(without alignments)
5032.400 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAHH.....NPQVASALTDLKEGLEMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Genesep_29Jan04.*

1: genesep1980s.*

2: genesep1990s.*

3: genesep2000s.*

4: genesep2001s.*

5: genesep2002s.*

6: genesep2003as.*

7: genesep2003bs.*

8: genesep2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 AAW98011	Hyperse
2	9448	100.0	1838	3 AAY71095	Erwinia a
3	9448	100.0	1838	3 AAY84856	A hyperse
4	9448	100.0	1838	4 AAU02878	Erwinia a
5	9448	100.0	1838	5 AAU16449	Erwinia a
6	2723	28.8	518	4 AAU04490	Erwinia a
7	2563	27.1	507	4 AAU04491	Erwinia a
8	1041	11.0	201	3 AAY69258	Amino aci
9	302	3.2	3041	6 ABM70325	Phototrab
10	295.5	3.1	2975	6 ABM68759	Phototrab
11	295	3.1	2951	6 ABM67717	Phototrab
12	280.5	3.0	4560	6 ABM67454	Phototrab
13	275	2.9	2514	3 AAY75097	Neisseria
14	275	2.9	2514	6 ABU06028	N. mening
15	275	2.9	2514	6 ABU38197	Protein e
16	274.5	2.9	9535	6 ABM73008	Staphyloc
17	274.5	2.9	10498	6 ABU19119	Pathogen
18	273.5	2.9	6713	6 ABU15887	Protein e
19	267.5	2.8	2659	5 AAY75489	S. aureus
20	262.5	2.8	5024	4 AAG82935	S. epider
21	262	2.8	6281	4 AAU37403	Staphyloc
22	261	2.8	2799	6 ABU37640	Protein e
23	258.5	2.7	6641	6 ABU42656	Protein e
24	257.5	2.7	10182	5 ABP38314	Staphyloc
25	257	2.7	5795	4 AAU37017	Staphyloc

26	256.5	2.7	2398	6 ABU42252	Protein e
27	253	2.7	2045	6 ABU46539	Protein e
28	253	2.7	2059	5 ABP25711	Streptoco
29	248.5	2.6	3158	4 AAU37018	Staphyloc
30	247.5	2.6	2344	4 AAU37120	Staphyloc
31	245	2.6	2599	3 AAY75098	Neisseria
32	245	2.6	2703	6 ABU38184	Protein e
33	244.5	2.6	3259	7 ADE56037	Human pro
34	244.5	2.6	3259	7 ADE56033	Staphyloc
35	242	2.6	2434	4 AAU34339	Protein e
36	241.5	2.6	1837	3 AAY85564	Human hom
37	241	2.6	1577	6 ABU41145	Protein e
38	240.5	2.5	2086	4 AAU34143	Staphyloc
39	240.5	2.5	5533	4 ABB65772	Drosophil
40	240.5	2.5	5560	4 ABB71160	Drosophil
41	239	2.5	2712	6 ABU39146	Protein e
42	238.5	2.5	2504	6 ADA34534	Acinetoba
43	238	2.5	2261	6 ABU18914	Pathogen
44	238	2.5	2271	6 ABM72734	Staphyloc
45	238	2.5	2283	6 ABP56876	Staphyloc

ALIGNMENTS

RESULT 1

AAW98011

ID AAW98011 standard; protein; 1838 AA.

XX AAW98011;

XX 21-JUN-1999 (first entry)

DE Hypersensitive response elicitor DspE.

XX Hypersensitive response elicitor; DspE; disease resistance;

KW insect resistance; biological control; transgenic plant.

XX Erwinia amylovora.

XX WO9907206-A1.

XX 18-FEB-1999.

XX 24-JUL-1998; 98WO-US015426.

XX 06-AUG-1997; 97US-0055105P.

XX (CORR) CORNELL RES FOUND INC.

XX Bogdanove AJ, Kim JF, Wei Z, Beer SV;

XX WPI; 1999-180362/15.

XX N-PSDB; AAX24810.

PT Nucleic acid encoding hypersensitive response-eliciting protein - used to improve growth of plants and impart resistance to disease and insects.

XX Claim 18; Page 51-56; 75pp; English.

XX This polypeptide comprises the 198 kDa hypersensitive response elicitor protein DspE of Erwinia amylovora. The nucleotide sequence of the dsp region of E. amylovora strain Ea321 was determined using subclones of pCP430. A two-gene operon was discovered comprising dspE (see AAX24810) and dspF (see AAX24811). The isolated dsp DNA molecules and encoded proteins can be used to impart disease resistance to plants, to enhance plant growth, and/or to control insects on plants. This is achieved by applying a hypersensitive response elicitor protein or polypeptide in a non-infectious form to plants or plant seeds, or by producing transgenic plants or plant seeds transformed with DNA encoding a hypersensitive response elicitor. Protection can be provided against a wide range of viruses, bacteria, fungi and insects, e.g. tobacco mosaic virus and tomato mosaic virus, Pseudomonas syringae, Xanthomonas campestris,

CC Fusarium oxysporum, Phytophthora infestans, armyworm, diamondback moth,
 CC etc. The method avoids use of infectious agents or polluting chemicals.
 CC Claimed transgenic plants are selected from alfalfa, rice, wheat, barley,
 CC rye, cotton, sunflower, peanut, corn, potato, bean, pea, chichory,
 CC lettuce, endive, cabbage, brussel sprout, sweet potato, beet, parsnip,
 CC turnip, cauliflower, broccoli, turnip, radish, spinach, onion, garlic,
 CC eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber,
 CC apple, pear, melon, citrus, strawberry, grape, raspberry, pineapple,
 CC soybean, tobacco, tomato, sorghum, sugarcane, Arabidopsis thaliana,
 CC Sainpaulia, petunia, pelargonium, poinsettia, chrysanthemum, carnation
 CC and zinnia
 XX
 SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 2; Length 1838;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSPQNAASLAAEGKRGKMPRIHQ 60
 DB 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSPQNAASLAAEGKRGKMPRIHQ 60
 QY 61 STAADGISAAHQKKSFSLRGCLGTTKFSRSAPQGGTTHSKGATLRDLARDGETQH 120
 DB 61 STAADGISAAHQKKSFSLRGCLGTTKFSRSAPQGGTTHSKGATLRDLARDGETQH 120
 QY 121 EAAAPDAALTRSGGVKRRNMDMAGRPVYKGGSGEDKVPYQKRHLNFGQMTLS 180
 DB 121 EAAAPDAALTRSGGVKRRNMDMAGRPVYKGGSGEDKVPYQKRHLNFGQMTLS 180
 QY 181 KMAHPASAGDRLQSPHPIPOSHHEIKEEPPVGTSTKATTAHADRVETIAQEDDDSEFQ 240
 DB 181 KMAHPASAGDRLQSPHPIPOSHHEIKEEPPVGTSTKATTAHADRVETIAQEDDDSEFQ 240
 QY 241 LHQORLARENPPOPPLGVATPI SARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
 DB 241 LHQORLARENPPOPPLGVATPI SARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
 QY 301 GAGVTPAVLTLDKGLQOLAPNDPAPLNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360
 DB 301 GAGVTPAVLTLDKGLQOLAPNDPAPLNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360
 QY 361 HLFDDIKSTATSVLHNSHPGETKGLAQAGTGSVSDGSKGKISLGSTQSHNKTMLS 420
 DB 361 HLFDDIKSTATSVLHNSHPGETKGLAQAGTGSVSDGSKGKISLGSTQSHNKTMLS 420
 QY 421 PGEAHRSLTGTWHPAGAAPOGESIRLHDDKIHILHPELGVQWQADKTHQSLSRAD 480
 DB 421 PGEAHRSLTGTWHPAGAAPOGESIRLHDDKIHILHPELGVQWQADKTHQSLSRAD 480
 QY 481 GKLYALKONRTQNLSDNKSSEKLVDKIKSYSDQGVQVAILTDTTPGRHKMSIMPDLAS 540
 DB 481 GKLYALKONRTQNLSDNKSSEKLVDKIKSYSDQGVQVAILTDTTPGRHKMSIMPDLAS 540
 QY 541 PESHISLSLHFAADHAGHGLHSELEAQSVALSHGRLVVDSEGLFSAIIPKQDGNEL 600
 DB 541 PESHISLSLHFAADHAGHGLHSELEAQSVALSHGRLVVDSEGLFSAIIPKQDGNEL 600
 QY 601 KMKAMPQHALDEHFGHDHQISGFFHDDHGLNALVKNFRQQAACPLGNDHOFHPGWNLT 660
 DB 601 KMKAMPQHALDEHFGHDHQISGFFHDDHGLNALVKNFRQQAACPLGNDHOFHPGWNLT 660
 QY 661 DALVIDNQGLHHTPEPHEIILDMHGLSLALQEGKLYFDQLTGKWTGAESDCKQLKKG 720
 DB 661 DALVIDNQGLHHTPEPHEIILDMHGLSLALQEGKLYFDQLTGKWTGAESDCKQLKKG 720
 QY 721 LDGAAYLLKDGVRKLNINQSTSSIKHGTENVFSLPHVENKPEPGDALQGNKDKQAAM 780
 DB 721 LDGAAYLLKDGVRKLNINQSTSSIKHGTENVFSLPHVENKPEPGDALQGNKDKQAAM 780
 QY 781 AVIGNVKYALTEKGDIRSFQIKPGTQQLERPAQTLREGISGELKDINHVDHKQNLALT 840
 DB 781 AVIGNVKYALTEKGDIRSFQIKPGTQQLERPAQTLREGISGELKDINHVDHKQNLALT 840

RESULT 2
 AAY71095
 ID AAY71095 standard; protein; 1838 AA.

QY 841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLSLDMGSHKPIATPFDGSOHLKAG 900
 DB 841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLSLDMGSHKPIATPFDGSOHLKAG 900
 QY 901 GWHAYAAPRGPIAVGTSGSTQVFNRLMOGVKGVIPGSLTVKLQAQGTGGMTCAGRKV 960
 DB 901 GWHAYAAPRGPIAVGTSGSTQVFNRLMOGVKGVIPGSLTVKLQAQGTGGMTCAGRKV 960
 QY 961 SSKFSERIIRAYAFNPTMSTPRPIKNAAYATQHOMQREGKLPLEYQGALIKOLDANVR 1020
 DB 961 SSKFSERIIRAYAFNPTMSTPRPIKNAAYATQHOMQREGKLPLEYQGALIKOLDANVR 1020
 QY 1021 HNAQPDLQSKLETLDLGEHGAELNDMKRFRDELFQSATRSVTVLQGHGVLSKNGEIN 1080
 DB 1021 HNAQPDLQSKLETLDLGEHGAELNDMKRFRDELFQSATRSVTVLQGHGVLSKNGEIN 1080
 QY 1081 SEFKPSFGKALVQSFNNRSGQDLKSLOOAHAHPSPAPSKLOSMGHFVSAGVDMHQ 1140
 DB 1081 SEFKPSFGKALVQSFNNRSGQDLKSLOOAHAHPSPAPSKLOSMGHFVSAGVDMHQ 1140
 QY 1141 KGEIPLGRORDPNDKTALTKSRLILDTVTIGELHELADKAKVSDHKPDADQIKOLRQOF 1200
 DB 1141 KGEIPLGRORDPNDKTALTKSRLILDTVTIGELHELADKAKVSDHKPDADQIKOLRQOF 1200
 QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTTTTLVLESQ 1260
 DB 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTTTTLVLESQ 1260
 QY 1261 SAEALAKLQNTLLSLDSGESMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
 DB 1261 SAEALAKLQNTLLSLDSGESMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
 QY 1321 LSFRTSGGLNVSGFDCGVGNIVATGHDVYPMYTKKTSGAGNADWLSAKHKTSPDL 1380
 DB 1321 LSFRTSGGLNVSGFDCGVGNIVATGHDVYPMYTKKTSGAGNADWLSAKHKTSPDL 1380
 QY 1381 RIGAAVSGTLQTLQNSLKFCLTEDELPGFIHGLTHGTLTPAELLQKGEHOMKQSKLT 1440
 DB 1381 RIGAAVSGTLQTLQNSLKFCLTEDELPGFIHGLTHGTLTPAELLQKGEHOMKQSKLT 1440
 QY 1441 FSVDTSANLDDLRAGINLNDGSKPVGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500
 DB 1441 FSVDTSANLDDLRAGINLNDGSKPVGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500
 QY 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPGVIPPFTSTNVSAALADNRTSQSI 1560
 DB 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPGVIPPFTSTNVSAALADNRTSQSI 1560
 QY 1561 SLELKRABPVTSDISELTSTLKGKFKDSATTMLAALKELDDAKPAEQHLIIQQHFSAK 1620
 DB 1561 SLELKRABPVTSDISELTSTLKGKFKDSATTMLAALKELDDAKPAEQHLIIQQHFSAK 1620
 QY 1621 DVVGDERYEAVNLKLVIRQQAADSHMELGASHSTYNNLSRINNDGIVELLHKHFD 1680
 DB 1621 DVVGDERYEAVNLKLVIRQQAADSHMELGASHSTYNNLSRINNDGIVELLHKHFD 1680
 QY 1681 AALPASSAKRLGEMWNPDPALKDIIKQLQSTPFSSASVSMELDGLREOTEKAILDGKVG 1740
 DB 1681 AALPASSAKRLGEMWNPDPALKDIIKQLQSTPFSSASVSMELDGLREOTEKAILDGKVG 1740
 QY 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSERNIGTINFKYG 1800
 DB 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSERNIGTINFKYG 1800
 QY 1801 ODQONTPRFTLEGGIAQANPQVASALTDLKKEGLEWKS 1838
 DB 1801 ODQONTPRFTLEGGIAQANPQVASALTDLKKEGLEWKS 1838

XX
AC
XX
DT
XX
DE
DE
XX
KW
KW
XX
OS
XX
XX
PN
XX
PD
XX
PF
XX
PR
XX
PA
XX
PI
XX
XX
DR
DR
XX
PT
PT
XX
PS
XX
SQ

AAAY71095;
08-SEP-2000 (first entry)
Erwinia amylovora hypersensitive response elicitor encoded by dspE gene.
Hypersensitive response elicitor; environmental stress resistance; plant;
pathogen; dapE gene.
Erwinia amylovora.
WO200028055-A2.
18-MAY-2000.
04-NOV-1999; 99WO-US026039.
05-NOV-1998; 98US-0107243P.
(EDEN-) EDEN BIOSCIENCE CORP.
Wei Z, Schading RL;
WPI; 2000-376566/32.
N-PSDB; AAD00670.
Application of a hypersensitive response elicitor protein to plants to
impart stress resistance.
Disclosure; Page 15-20; 84pp; English.
The patent discloses a method to impart stress resistance to plants by
applying a hypersensitive response elicitor in a non-infectious form to a
plant or seed. The present sequence is a hypersensitive response elicitor
encoded by dspE gene from Erwinia amylovora. The protein elicits plant
pathogen's hypersensitive response and is used to impart stress
resistance to plants
Sequence 1838 AA;
Query Match 100.0%; Score 9448; DB 3; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOQSSSSSPQNAASLAABGKRGKMPRIHQ 60
DB 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOQSSSSSPQNAASLAABGKRGKMPRIHQ 60
QY 61 STAADGISAHOQKFSLRGCLGTCKFSRPOGQGTTHSKGATLRDLARDGETQH 120
DB 61 STAADGISAHOQKFSLRGCLGTCKFSRPOGQGTTHSKGATLRDLARDGETQH 120
QY 121 EAAAPDAARLTRSGVKRRNMDMAGRPMVKGSGEDKVPQTKRHLNPNFQMRQMTLS 180
DB 121 EAAAPDAARLTRSGVKRRNMDMAGRPMVKGSGEDKVPQTKRHLNPNFQMRQMTLS 180
QY 181 KWAHPASAGDRLOHSPPHIPGSHHEIKKEPVGSTSKATTAHADRVELIAEDDDSEFQ 240
DB 181 KWAHPASAGDRLOHSPPHIPGSHHEIKKEPVGSTSKATTAHADRVELIAEDDDSEFQ 240
QY 241 LHOQLARERENPPPKLGWATPISARFQPLTAVAESVLEGDTTOSPLKPOSMLKGS 300
DB 241 LHOQLARERENPPPKLGWATPISARFQPLTAVAESVLEGDTTOSPLKPOSMLKGS 300
QY 301 GAGVTPLAVTLDKGKLQAPDNPALNTLLKOTLGKDTQHYLAHASSDGSQHLLDNKG 360
DB 301 GAGVTPLAVTLDKGKLQAPDNPALNTLLKOTLGKDTQHYLAHASSDGSQHLLDNKG 360
QY 361 HLFDIKSTATSVYLHNSHPGIEIKGLAQAGTGSVSDGSKGKSLGSGTQSHNKTMLSQ 420
DB 361 HLFDIKSTATSVYLHNSHPGIEIKGLAQAGTGSVSDGSKGKSLGSGTQSHNKTMLSQ 420

QY 421 PGEAHRSLLTGTWQHPAGAARPOGESIRLHDDKIHLHPELGWQSQADKDTQSLSROAD 480
DB 421 PGEAHRSLLTGTWQHPAGAARPOGESIRLHDDKIHLHPELGWQSQADKDTQSLSROAD 480
QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQGVAILDTPGRHKMSIMPSLDAS 540
DB 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQGVAILDTPGRHKMSIMPSLDAS 540
QY 541 PESHISLSLHFADAHQGLLHGKSELEAQAQVAISHGRVLVADSEGLFSAAPKQGDGNE 600
DB 541 PESHISLSLHFADAHQGLLHGKSELEAQAQVAISHGRVLVADSEGLFSAAPKQGDGNE 600
QY 601 KMKAMPQHALDEHFHGHQISGFFHDDHQLNALVKNFRQHQHACPLGNHQPFGWNL 660
DB 601 KMKAMPQHALDEHFHGHQISGFFHDDHQLNALVKNFRQHQHACPLGNHQPFGWNL 660
QY 661 DALVIDNQLGLHHTNPPEHILDMGHLSLALQEGKLIHYFDLTCKWTGAESCKQLKKG 720
DB 661 DALVIDNQLGLHHTNPPEHILDMGHLSLALQEGKLIHYFDLTCKWTGAESCKQLKKG 720
QY 721 LDGAAYLLKDGGEVKRLNINQSTSSIKHGTENVSFLPHVRNKPFGDALQGLNKKDAQAM 780
DB 721 LDGAAYLLKDGGEVKRLNINQSTSSIKHGTENVSFLPHVRNKPFGDALQGLNKKDAQAM 780
QY 781 AVIGVKNYLALTEKGDIRSFIKPGTQQLERPAQTLSREGISGELKDIIHVDHQNLYALT 840
DB 781 AVIGVKNYLALTEKGDIRSFIKPGTQQLERPAQTLSREGISGELKDIIHVDHQNLYALT 840
QY 841 HEGEVFHQPREAWONGAESSEWHKLALPOSEKLSLDMSHEHKPIATFEDGSHQKAG 900
DB 841 HEGEVFHQPREAWONGAESSEWHKLALPOSEKLSLDMSHEHKPIATFEDGSHQKAG 900
QY 901 GWIYAAPRGPLAVGTSGSQTGFNRLMOGVKGVIPGSGLTVKLSAQTCGTMGAERKV 960
DB 901 GWIYAAPRGPLAVGTSGSQTGFNRLMOGVKGVIPGSGLTVKLSAQTCGTMGAERKV 960
QY 961 SSKFSERIRAYANPNWSTPRPIKNAAYATQHGQWREGKLPKLYEMQALIKQJDAHNVR 1020
DB 961 SSKFSERIRAYANPNWSTPRPIKNAAYATQHGQWREGKLPKLYEMQALIKQJDAHNVR 1020
QY 1021 HNAPODLOSKLETLDLGEHGAELLNDMKRPRELEQASRSTVTLGQHVLSKNGEIN 1080
DB 1021 HNAPODLOSKLETLDLGEHGAELLNDMKRPRELEQASRSTVTLGQHVLSKNGEIN 1080
QY 1081 SEFKPSGKALVQSFNVNRSQDLKSLQQAQVATPPSAESKLSQMLGHFVSAGVDMHQ 1140
DB 1081 SEFKPSGKALVQSFNVNRSQDLKSLQQAQVATPPSAESKLSQMLGHFVSAGVDMHQ 1140
QY 1141 KGEIPLGRQDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLROOF 1200
DB 1141 KGEIPLGRQDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLROOF 1200
QY 1201 DTUREKYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKXBBHGVNLTTRTVLESQ 1260
DB 1201 DTUREKYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKXBBHGVNLTTRTVLESQ 1260
QY 1261 SAEIAXKLLKNTLLSLDSESMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
DB 1261 SAEIAXKLLKNTLLSLDSESMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
QY 1321 LSFRTSGGLNVSPRGDGVSGNIMVATGHVMPYMTGKTSAGNASDWLSAKHKISPD 1380
DB 1321 LSFRTSGGLNVSPRGDGVSGNIMVATGHVMPYMTGKTSAGNASDWLSAKHKISPD 1380
QY 1381 RIGAAVSGTLOGLTQNSLKPKLDELPFGTHGLTHGLTTPAELLKXGIEHOMKQGSKLT 1440
DB 1381 RIGAAVSGTLOGLTQNSLKPKLDELPFGTHGLTHGLTTPAELLKXGIEHOMKQGSKLT 1440
QY 1441 FSVDTSANLDRAGINLNDGSKPVGTVARVAGLSASANLAAGSRERSTTSGFGSTTS 1500
DB 1441 FSVDTSANLDRAGINLNDGSKPVGTVARVAGLSASANLAAGSRERSTTSGFGSTTS 1500
QY 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAPTSTNVSAALADNRTSOSI 1560

1501 ASNNRPTFLNGVAGANLTAALGVHSSTHEGPVGIFFAFTSTNSAALADNRTSQSI 1560
1561 SLELKRAEPTVNDISLSTLGLGHFKDSATTWMLAALKELDDAKPAEQHLHILQOHFSK 1620
1561 SLELKRAEPTVNDISLSTLGLGHFKDSATTWMLAALKELDDAKPAEQHLHILQOHFSK 1620
1621 DVVGDERYEAVRNLLKLVIRQQAADSHMELGSAHSHTYNNLSRINNDGIVELLHKHFD 1680
1621 DVVGDERYEAVRNLLKLVIRQQAADSHMELGSAHSHTYNNLSRINNDGIVELLHKHFD 1680
1681 AALPASSAKELGEMMNDPALKDILKLOLSTPSSASVSMELKDLGRLQTEKAILDGKVG 1740
1681 AALPASSAKELGEMMNDPALKDILKLOLSTPSSASVSMELKDLGRLQTEKAILDGKVG 1740
1741 REEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
1741 REEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
1801 QDQNTPRFTLEGGIAQANPQVASALTDILKKEGLEWKS 1838
1801 QDQNTPRFTLEGGIAQANPQVASALTDILKKEGLEWKS 1838

RESULT 3

AA84856

ID AA84856 standard; protein; 1838 AA.

AC AC

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

CC melon, citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia, petunia,
CC pelargonium, poinsettia, chrysanthemum, carnation or zinnia
XX
SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 3; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQCGSSSSSPQNAASLAIEGKRGKMPRIHQ 60
Db 1 MELKSLGTEHKAHVHTAAHNPVGHVALQCGSSSSSPQNAASLAIEGKRGKMPRIHQ 60
QY 61 STAADGISAHQOKKFSLRGCLGTTKFFRSAPQGGPQGTTHSGATLRDLARDGGTQH 120
Db 61 STAADGISAHQOKKFSLRGCLGTTKFFRSAPQGGPQGTTHSGATLRDLARDGGTQH 120
QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVPTQKRLHNNFQGMQRTMLS 180
Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVPTQKRLHNNFQGMQRTMLS 180
QY 181 KMAHPASANAGDRLQHSPPHIPSHEHEIKKEPVGTSKATTAHADRVETAEQDDSEFQ 240
Db 181 KMAHPASANAGDRLQHSPPHIPSHEHEIKKEPVGTSKATTAHADRVETAEQDDSEFQ 240
QY 241 LHOQRLARENPPOPPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKXPQSMKGS 300
Db 241 LHOQRLARENPPOPPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKXPQSMKGS 300
QY 301 GAGVTPPLAVTLQKQLQAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360
Db 301 GAGVTPPLAVTLQKQLQAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360
QY 361 HLFDTIKSTATSYVLHNSHPGKIGKLAQAGTSVSDGSKGKISLGSTQSHNKMTLSQ 420
Db 361 HLFDTIKSTATSYVLHNSHPGKIGKLAQAGTSVSDGSKGKISLGSTQSHNKMTLSQ 420
QY 421 PGEAHRSLTIGIWOHPAGAAARPOGESIRLHDDKIHLHPGLVWQSAADKDTLSQSRQAD 480
Db 421 PGEAHRSLTIGIWOHPAGAAARPOGESIRLHDDKIHLHPGLVWQSAADKDTLSQSRQAD 480
QY 481 GKLYALKONRTLQNLSDNKSSEKLVDTKISYSDQGVAILTDTTPGRHKMSIMPSLDAS 540
Db 481 GKLYALKONRTLQNLSDNKSSEKLVDTKISYSDQGVAILTDTTPGRHKMSIMPSLDAS 540
QY 541 PESHISLSLHFADAHQGLLHGKSELAQSVASHGRLVVDSEGLKLSAAIPKQGGNGL 600
Db 541 PESHISLSLHFADAHQGLLHGKSELAQSVASHGRLVVDSEGLKLSAAIPKQGGNGL 600
QY 601 KMKAMPQHALDEHFGHDHQSIFGFFHDDHQLNALVKNFRQOQHACPLGNDHQHPQWNL 660
Db 601 KMKAMPQHALDEHFGHDHQSIFGFFHDDHQLNALVKNFRQOQHACPLGNDHQHPQWNL 660
QY 661 DALVTDNQLGLHHTNPEPHEILDMGHLSLALQEGKLYHFDQLTGKWTGAESDCKOLKKG 720
Db 661 DALVTDNQLGLHHTNPEPHEILDMGHLSLALQEGKLYHFDQLTGKWTGAESDCKOLKKG 720
QY 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPDPGALQGLNKKDAQAM 780
Db 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPDPGALQGLNKKDAQAM 780
QY 781 AVIGVKNYLATEKGDIRSFOIKPQTQQLERPAQTLREGISGELKD IHVDHKQNLALT 840
Db 781 AVIGVKNYLATEKGDIRSFOIKPQTQQLERPAQTLREGISGELKD IHVDHKQNLALT 840
QY 841 HEGEVFHQPRAWQNGAESSSWHKIALPQSESKLSLMSHEHKPIATPDSGSHOLKAG 900
Db 841 HEGEVFHQPRAWQNGAESSSWHKIALPQSESKLSLMSHEHKPIATPDSGSHOLKAG 900
QY 901 GWHVAAAPERGPLAVGTSGSTQVFNRLMQGVKGVIPGSLTVKLSAQTGGMTGAEGRKV 960
Db 901 GWHVAAAPERGPLAVGTSGSTQVFNRLMQGVKGVIPGSLTVKLSAQTGGMTGAEGRKV 960

QY 961 SSKFSRIRAYANPTMSTPRPIKNAAYATQHQMGREGKPLVEMOGALIKOLDAHNVR 1020

Db 961 SSKFSRIRAYANPTMSTPRPIKNAAYATQHQMGREGKPLVEMOGALIKOLDAHNVR 1020

QY 1021 HNAPODQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGQHQGVLSNGEIN 1080

Db 1021 HNAPODQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGQHQGVLSNGEIN 1080

QY 1081 SEFKPSGKALVOSFNVRSGQDLSKLSLOQAVHATPPSAESKLSMLGHFVSAGVMSHQ 1140

Db 1081 SEFKPSGKALVOSFNVRSGQDLSKLSLOQAVHATPPSAESKLSMLGHFVSAGVMSHQ 1140

QY 1141 KGRIPLGRQDPNDKTKALTSRLILDTVTIGELHELADKAKLYSDHKFPDADQIKQLRQOF 1200

Db 1141 KGRIPLGRQDPNDKTKALTSRLILDTVTIGELHELADKAKLYSDHKFPDADQIKQLRQOF 1200

QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQ 1260

Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQ 1260

QY 1261 SAEKAKLNTLISLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAIN 1320

Db 1261 SAEKAKLNTLISLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAIN 1320

QY 1321 LSPSRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKKTSGNAGSDWLSAKHKISPD 1380

Db 1321 LSPSRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKKTSGNAGSDWLSAKHKISPD 1380

QY 1381 RIGAAVSGTLQGTLONSLKPFLTEDELPFGIHTGLTPAELLQKGIHQMKGSKLT 1440

Db 1381 RIGAAVSGTLQGTLONSLKPFLTEDELPFGIHTGLTPAELLQKGIHQMKGSKLT 1440

QY 1441 FSVDTSANLDRAGINLNDGSPNGVTARVSGLSANLAAGSRERSTTSGQFGSTTS 1500

Db 1441 FSVDTSANLDRAGINLNDGSPNGVTARVSGLSANLAAGSRERSTTSGQFGSTTS 1500

QY 1501 ASNNRPTFLNGVAGANLTAALGVHSGTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Db 1501 ASNNRPTFLNGVAGANLTAALGVHSGTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

QY 1561 SLEKRAEPTVNSIDELSTLTKGHFKDSATTKMLAALKELDDAKPAEQLHILQQHPSAK 1620

Db 1561 SLEKRAEPTVNSIDELSTLTKGHFKDSATTKMLAALKELDDAKPAEQLHILQQHPSAK 1620

QY 1621 DVVGDREYEAURNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Db 1621 DVVGDREYEAURNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

QY 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPSSASVSMELKDGLREOTEKAILDGKVG 1740

Db 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPSSASVSMELKDGLREOTEKAILDGKVG 1740

QY 1741 REEVGVLPQDRNNLRVKSVSQSVKSEGFNTPALLLGTSNAAAMERNIGTINPKYG 1800

Db 1741 REEVGVLPQDRNNLRVKSVSQSVKSEGFNTPALLLGTSNAAAMERNIGTINPKYG 1800

QY 1801 QDQNTPRFTLEGIAQANPOVASALTDLAKKEGLEMS 1838

Db 1801 QDQNTPRFTLEGIAQANPOVASALTDLAKKEGLEMS 1838

RESULT 4

AAU02878

ID AAU02878 standard; protein; 1838 AA.

XX

AC AAU02878;

XX

DT 23-OCT-2001 (first entry)

XX

DE Erwinia amylovora disease-specific region E (dspE) polypeptide.

XX

XX Disease-specific region E; dspE; insect control; disease resistance;

hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic; cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr; petunia; chrysanthemum; carnation; transgenic plant; avirulence locus; pathogenicity; fireblight.

XX Erwinia amylovora.

XX OS US6228644-B1.

XX PN 08-MAY-2001.

XX PD 22-JUL-1998; 98US-00120663.

XX PF 06-AUG-1997; 97US-0055106P.

XX PR (CORR) CORNELL RES FOUND INC.

XX PA Bogdanove AJ, Kim JF, Wei Z, Beer SV;

XX PI WPI; 2001-327491/34.

XX DR N-PSDB; AAS05199.

XX PS Polynucleotides encoding hypersensitive response eliciting proteins or polypeptides useful for imparting disease resistance to plants, to enhance plant growth, and/or to control insects on plants.

XX CC Claim 1; Col 9-18; 37pp; English.

XX CC The sequence represents an Erwinia amylovora disease-specific region E (dspE) polypeptide, which elicits a hypersensitive response in plants. The dspE operon functions as an avirulence (avr) locus and the dspE protein is required for pathogenicity of Erwinia amylovora. DspE is also responsible for causing fireblight in certain plants. The protein and its associated nucleic acid can be applied in a non-infectious form to plants or plant seeds to impart disease resistance, to enhance plant growth, and/or to control insects on plants. Alternatively, plant cells may be transformed with the DNA to form transgenic plants with the same properties. The method can be utilised to treat a wide variety of plants and seeds, including crop plants such as rice, wheat, corn, cabbage, cauliflower, garlic, cucumber, apple, grape, tobacco and sugarcane, and ornamental plants such as Arabidopsis thaliana, petunia, chrysanthemum and carnation

xx SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 4; Length 1838;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVVALQGGSSSSPQNAASLAARGNKRGKMPRIHQ 60

Db 1 MELKSLGTEHKAHVHTAAHNPVGHVVALQGGSSSSPQNAASLAARGNKRGKMPRIHQ 60

QY 61 STAADGISAAHQKQKFSLRGCLGTTKFSRSAPQGGPQTHSGATLRLDARDGGTQH 120

Db 61 STAADGISAAHQKQKFSLRGCLGTTKFSRSAPQGGPQTHSGATLRLDARDGGTQH 120

QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVVKGGSGDKVPTQOKRHLNFGOMQRTMLS 180

Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVVKGGSGDKVPTQOKRHLNFGOMQRTMLS 180

QY 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPVSGTSKATTAAHADRVEIAQEDDDSEFQ 240

Db 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPVSGTSKATTAAHADRVEIAQEDDDSEFQ 240

QY 241 LHOORLARENPPOPKLGVATPI SARFOPKLTAVAESVLEGTDTTQSPKLPQSMKLG 300

Db 241 LHOORLARENPPOPKLGVATPI SARFOPKLTAVAESVLEGTDTTQSPKLPQSMKLG 300

QY 301 GAGVTPLAVTLDDKGLQALPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360

Db 301 GAGVTPLAVTLDDKGLQALPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360

361 QY HLFIDIKSTATSVLHNSHPGIBKGLAQAGTSGSVSDGSGKISLGSGTQSHNKMTLSQ 420
Db |||||
361 HLFIDIKSTATSVLHNSHPGIBKGLAQAGTSGSVSDGSGKISLGSGTQSHNKMTLSQ 420
QY |||||
421 PGEAHRSLTGTWHPAGAAAPQGESIRLHDDKIHILHPELGVWQADKTHQSQRQAD 480
Db |||||
421 PGEAHRSLTGTWHPAGAAAPQGESIRLHDDKIHILHPELGVWQADKTHQSQRQAD 480
QY |||||
481 GKLYALKDNRTLQNTSDNKSSEKLVDKIKSYSDVQGOVAIITDTPGRHKMSIMPSLDAS 540
Db |||||
481 GKLYALKDNRTLQNTSDNKSSEKLVDKIKSYSDVQGOVAIITDTPGRHKMSIMPSLDAS 540
QY |||||
541 PESHISLSLHFAHAGHGLHGKSELEAQAQVAISHGRLVWADSEGKLFSAAIKQGDGDEL 600
Db |||||
541 PESHISLSLHFAHAGHGLHGKSELEAQAQVAISHGRLVWADSEGKLFSAAIKQGDGDEL 600
QY |||||
601 KMKAMPQHALLDHFHGDHIOISGFHDDHGLNALVKNFRQOAHACPLGNDHOFHPGWNLT 660
Db |||||
601 KMKAMPQHALLDHFHGDHIOISGFHDDHGLNALVKNFRQOAHACPLGNDHOFHPGWNLT 660
QY |||||
661 DALVIDNQLGLHHTPEPEHEIIDMGHLSLALQEGKLYHFDQITKGTGAESDCKQLKKG 720
Db |||||
661 DALVIDNQLGLHHTPEPEHEIIDMGHLSLALQEGKLYHFDQITKGTGAESDCKQLKKG 720
QY |||||
721 LDGAAYLLKDGVEKRLININQSTSSIKHGTENVFSLPHVRNKPPEPGDALQGLNKDDKAQAM 780
Db |||||
721 LDGAAYLLKDGVEKRLININQSTSSIKHGTENVFSLPHVRNKPPEPGDALQGLNKDDKAQAM 780
QY |||||
781 AVIGUNKYIALTEKGIIRFQIKPGTQQLERPAQTLRSREGISGELDIHVDHKQNIYALT 840
Db |||||
781 AVIGUNKYIALTEKGIIRFQIKPGTQQLERPAQTLRSREGISGELDIHVDHKQNIYALT 840
QY |||||
841 HGEVFPHPQPREAWQNGAESSSHHKLLALPOSESKLSLDMSHHEKPIATFEDSGSQHOLKAG 900
Db |||||
841 HGEVFPHPQPREAWQNGAESSSHHKLLALPOSESKLSLDMSHHEKPIATFEDSGSQHOLKAG 900
QY |||||
901 GWEHAYAAPRGPLAVGTSGSTVFNRLMQGVGKVIPIGSLTVKLJSAQTGGMGTGAERKVK 960
Db |||||
901 GWEHAYAAPRGPLAVGTSGSTVFNRLMQGVGKVIPIGSLTVKLJSAQTGGMGTGAERKVK 960
QY |||||
961 SKKFSERIAYAPNPTMSTPRPKNAAYATQGWQREGKLPLEYMOGALLIKQLDAHNV 1020
Db |||||
961 SKKFSERIAYAPNPTMSTPRPKNAAYATQGWQREGKLPLEYMOGALLIKQLDAHNV 1020
QY |||||
1021 HNAPODLOKLETLDLGHEGABELNDMKRFRDELEQSATRSVTVLGQHGVLKNGEIN 1080
Db |||||
1021 HNAPODLOKLETLDLGHEGABELNDMKRFRDELEQSATRSVTVLGQHGVLKNGEIN 1080
QY |||||
1081 SEFKPSPGKALVQSFVNVRSGQDLKSLOQAHAATPPSAESKLSQMLGHFVSAGVDMSHQ 1140
Db |||||
1081 SEFKPSPGKALVQSFVNVRSGQDLKSLOQAHAATPPSAESKLSQMLGHFVSAGVDMSHQ 1140
QY |||||
1141 KGEIPLGRORDNDKLTALTKSLIILDTVTIGELHELADKAKLVSDHKPDADQIKQLROOF 1200
Db |||||
1141 KGEIPLGRORDNDKLTALTKSLIILDTVTIGELHELADKAKLVSDHKPDADQIKQLROOF 1200
QY |||||
1201 DTLREKRYESNPVXYHTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260
Db |||||
1201 DTLREKRYESNPVXYHTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260
QY |||||
1261 SHELAKLNTLLSDSGSMGFSRSYGGGVSTVFVPTLSKVPVPVPIPCAGITLDRAYN 1320
Db |||||
1261 SHELAKLNTLLSDSGSMGFSRSYGGGVSTVFVPTLSKVPVPVPIPCAGITLDRAYN 1320
QY |||||
1321 LSFRTSGGLNVSPFRDGGVGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380
Db |||||
1321 LSFRTSGGLNVSPFRDGGVGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380
QY |||||
1381 RIGAAVSGTLQGTQNSLAFKLTDELPGFIHGLTHGTLTPAELQKGIHEQMKGSKLT 1440
Db |||||
1381 RIGAAVSGTLQGTQNSLAFKLTDELPGFIHGLTHGTLTPAELQKGIHEQMKGSKLT 1440
QY |||||
1441 FSVDTISANLDRAGINLNEGSKPNGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500

1441 Db FSVDTISANLDRAGINLNEGSKPNGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500
QY |||||
1501 ASNNRPTFLNGVAGANLTAALGVVAHSSTHEGKPVGIFPAFTSTNVSAALADNRTSQSI 1560
Db |||||
1501 ASNNRPTFLNGVAGANLTAALGVVAHSSTHEGKPVGIFPAFTSTNVSAALADNRTSQSI 1560
QY |||||
1561 SLELKRAEPTVNDISELTSTLTKGPKDSATTKMLAALKELDDAKPAEOLHILQOHFSAK 1620
Db |||||
1561 SLELKRAEPTVNDISELTSTLTKGPKDSATTKMLAALKELDDAKPAEOLHILQOHFSAK 1620
QY |||||
1621 DVVGDERYEAVRNLKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLKHFD 1680
Db |||||
1621 DVVGDERYEAVRNLKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLKHFD 1680
QY |||||
1681 AALPASSAKRLGEMMNDPALKDIIKQLOSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
Db |||||
1681 AALPASSAKRLGEMMNDPALKDIIKQLOSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
QY |||||
1741 REEVGVLFODRNLVRKVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
Db |||||
1741 REEVGVLFODRNLVRKVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
QY |||||
1801 QDQNTPRRTLEGGIAQANPQVASALTDLKKEGLEWKS 1838
Db |||||
1801 QDQNTPRRTLEGGIAQANPQVASALTDLKKEGLEWKS 1838
RESULT 5
AAE16449
ID AAE16449 standard; protein; 1838 AA.
XX
AC AAE16449;
XX
DT 09-APR-2002 (first entry)
XX
DE Erwinia amylovora hypersensitive response elicitor protein, dspE.
XX
KW Hypersensitive response elicitor protein; plant growth; fruit coloration;
KW disease resistance; stress resistance; phytotoxin; insect infection;
KW plant maturation; dspE protein.
XX
OS Erwinia amylovora.
XX
PN WO200198501-A2.
XX
PD 27-DEC-2001.
XX
PF 12-JUN-2001; 2001WO-US018820.
XX
PR 16-JUN-2000; 2000US-0212211P.
XX
PA (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI Fan H, Wei Z;
XX
DR WPI; 2002-122282/16.
DR N-PSDB; AAD27017.
XX
PT New hypersensitive response elicitor proteins comprising spaced apart
PT domains having an acidic portion linked to an alpha-helix, useful for
PT imparting disease or stress resistance, controlling insects or enhancing
PT plant growth.
XX
PS Disclosure; Page 18-23; 99pp; English.
XX
CC The patent discloses hypersensitive response elicitor proteins and
CC nucleotides encoding such proteins. Hypersensitive response elicitor
CC proteins comprise an isolated pair or more of spaced apart domains, each
CC comprising an acidic portion linked to an alpha-helix and capable of
CC eliciting a hypersensitive response in plants. Sequences of the invention
CC are used to impart disease resistance to plants, to enhance plant growth,
CC to control insects and/or to impart stress resistance to plants which

CC includes resistance to environmental stresses such as climate, air
 CC pollution, chemical and nutritional stress. The method of imparting
 CC disease resistance has the potential for treating previously untreatable
 CC diseases, treating diseases systemically and avoiding the use of
 CC infectious agents or environmentally harmful materials. Hyper- sensitive
 CC response elicitor sequences are used to enhance plant growth which
 CC encompasses greater yield, increased in quantity of seeds produced,
 CC percentage of seeds germinated, plant size and biomass, bigger fruits,
 CC earlier fruit coloration and plant maturation. They are also used for
 CC insect control which encompasses preventing direct insect damage to plant
 CC by feeding injury, interfering with insect larval feeding on the plants,
 CC preventing insects from colonising host plants and releasing phytotoxins.
 CC Sequences of the invention also prevent subsequent disease damage to
 CC plants resulting from insect infection. The present sequence is Erwinia
 CC amylovora hypersensitive response elicitor protein, dspE
 XX
 SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 5; Length 1838;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELKSLGTEHKAHVTAHNPVGHVVALQQSSSSSPQNAASLAAGKNGKMPRIHQP 60
 Db 1 MELKSLGTEHKAHVTAHNPVGHVVALQQSSSSSPQNAASLAAGKNGKMPRIHQP 60

Qy 61 STAADGISAHQKPSFLRGCLGTKFPSAPOQPGTTHSKGATLRDLARDGETQH 120
 Db 61 STAADGISAHQKPSFLRGCLGTKFPSAPOQPGTTHSKGATLRDLARDGETQH 120

Qy 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVMKGGSGEDKVPQQKRRHQLNFPQMRQTMLS 180
 Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVMKGGSGEDKVPQQKRRHQLNFPQMRQTMLS 180

Qy 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKPEPVGSGTSKATTAHADRVIEIAQEDDDSEFQ 240
 Db 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKPEPVGSGTSKATTAHADRVIEIAQEDDDSEFQ 240

Qy 241 LHQQLARERENPPQPKLGAVATPISARFQPKLTAVAESVLEGGTDTQSPKLPKPSMLKGS 300
 Db 241 LHQQLARERENPPQPKLGAVATPISARFQPKLTAVAESVLEGGTDTQSPKLPKPSMLKGS 300

Qy 301 GAGVTPPLAVTLTDKGLQALPDPNPPALNTLLKQTGLKDTQHYLAHASSDGSQHLLDNKG 360
 Db 301 GAGVTPPLAVTLTDKGLQALPDPNPPALNTLLKQTGLKDTQHYLAHASSDGSQHLLDNKG 360

Qy 361 HLFDIKSTATSYSLVHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTQSHNKTMLSQ 420
 Db 361 HLFDIKSTATSYSLVHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTQSHNKTMLSQ 420

Qy 421 PGEAHSLLTGIWHPAGARPQGESIRLHDDKTHILHPELVGWSADKDTHSQLSRQAD 480
 Db 421 PGEAHSLLTGIWHPAGARPQGESIRLHDDKTHILHPELVGWSADKDTHSQLSRQAD 480

Qy 481 GKLYALKDNRTLQNLSDNKSEKLVDKISKYSVDQRGQVAILTDPGRHKMSIMPSLDAS 540
 Db 481 GKLYALKDNRTLQNLSDNKSEKLVDKISKYSVDQRGQVAILTDPGRHKMSIMPSLDAS 540

Qy 541 PESHISLSLHFADAHQGLLHGKSLBAQSVASHGRVLVADSEGLFSAALPKQGDGNEL 600
 Db 541 PESHISLSLHFADAHQGLLHGKSLBAQSVASHGRVLVADSEGLFSAALPKQGDGNEL 600

Qy 601 KMKAMPQHALDEHFGHDHIOISGFFHDDHGQLNALVKNNFRQHCACPLGNDHQFHPGNLIT 660
 Db 601 KMKAMPQHALDEHFGHDHIOISGFFHDDHGQLNALVKNNFRQHCACPLGNDHQFHPGNLIT 660

Qy 661 DALVIDNQLGLHHTNPPEHETLDMGHLGSLAQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
 Db 661 DALVIDNQLGLHHTNPPEHETLDMGHLGSLAQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy 721 LDGAAYLLKQGEVRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
 Db 721 LDGAAYLLKQGEVRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780

Qy 781 AVIGWNKYLALTEKGDIRSFOIKPGTQOLRPAOTLSREGISGELKDIIHVDHKQNLAYLT 840
 Db 781 AVIGWNKYLALTEKGDIRSFOIKPGTQOLRPAOTLSREGISGELKDIIHVDHKQNLAYLT 840

Qy 841 HEGEVFHQPREAMONGAESSWHKLALPQSESKLSIDMSHEHKPIATFEDGSHQLKAG 900
 Db 841 HEGEVFHQPREAMONGAESSWHKLALPQSESKLSIDMSHEHKPIATFEDGSHQLKAG 900

Qy 901 GWHAYAAPRGPLAVGTSGSQTENRLMQGVKQVIPSGLITVTKLSAQTCGWTGAEGEKV 960
 Db 901 GWHAYAAPRGPLAVGTSGSQTENRLMQGVKQVIPSGLITVTKLSAQTCGWTGAEGEKV 960

Qy 961 SSKFESERIRAYAFNPTMTSTPRPIKNAAYATQHGQWQREGKPLKPLYEMOGALIKQLDAHNR 1020
 Db 961 SSKFESERIRAYAFNPTMTSTPRPIKNAAYATQHGQWQREGKPLKPLYEMOGALIKQLDAHNR 1020

Qy 1021 HNAPOPDLQSKLETLDLGEHGAELNDMKFRDELEQSATRSVTVLGQHOGVLKNSGEIN 1080
 Db 1021 HNAPOPDLQSKLETLDLGEHGAELNDMKFRDELEQSATRSVTVLGQHOGVLKNSGEIN 1080

Qy 1081 SEFKPSGKALVQSFNVRSGODLSKSLQOAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140
 Db 1081 SEFKPSGKALVQSFNVRSGODLSKSLQOAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140

Qy 1141 KGEIPLGRQDPNDKTALTCSRILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1200
 Db 1141 KGEIPLGRQDPNDKTALTCSRILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1200

Qy 1201 DTLEKKEYESNPVKHYTDMGTTHNKALEANYDAVKAFINAPKKBHGVNLTTRTVLESQ 1260
 Db 1201 DTLEKKEYESNPVKHYTDMGTTHNKALEANYDAVKAFINAPKKBHGVNLTTRTVLESQ 1260

Qy 1261 SAEIAKLLKNTLLSLDSGESMSFSRSGGVSTVFVPTLSKKVPVPVPGAGITLDRAYN 1320
 Db 1261 SAEIAKLLKNTLLSLDSGESMSFSRSGGVSTVFVPTLSKKVPVPVPGAGITLDRAYN 1320

Qy 1321 LSFSTRSGGLNVSRGDRGVSGNIMVATGHDVMPYMTGKKTSAAGNSDWLSAKHKISPD 1380
 Db 1321 LSFSTRSGGLNVSRGDRGVSGNIMVATGHDVMPYMTGKKTSAAGNSDWLSAKHKISPD 1380

Qy 1381 RIGAAVSGTLQGTTLQNSLKFKLTDELPFGTHGLTHGTLTPAEILLQKQIEHOMKQSGKLT 1440
 Db 1381 RIGAAVSGTLQGTTLQNSLKFKLTDELPFGTHGLTHGTLTPAEILLQKQIEHOMKQSGKLT 1440

Qy 1441 FSDVTSANLDIRAGINLNEGSKPENGVTARVSAGLSASANLAAGSRERSTTSGFGSTTS 1500
 Db 1441 FSDVTSANLDIRAGINLNEGSKPENGVTARVSAGLSASANLAAGSRERSTTSGFGSTTS 1500

Qy 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSOSI 1560
 Db 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSOSI 1560

Qy 1561 SLELKRAEPTVNSDISELTSTLGHFKFDKSDATTKMLAALKELDDAKPAEQLHILQHFSAK 1620
 Db 1561 SLELKRAEPTVNSDISELTSTLGHFKFDKSDATTKMLAALKELDDAKPAEQLHILQHFSAK 1620

Qy 1621 DVVGDREYEAARNLKLVIQQAADSHMELGSAHSTTYNNLSRINNDGIVELLHKHFD 1680
 Db 1621 DVVGDREYEAARNLKLVIQQAADSHMELGSAHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy 1681 AALPASSAKSLGEMNNNDPALKDIKOLQSTPSSASVSMELKDLREQTEKALDGVK 1740
 Db 1681 AALPASSAKSLGEMNNNDPALKDIKOLQSTPSSASVSMELKDLREQTEKALDGVK 1740

Qy 1741 REEVGVLFQDRNNLRVKSVSVSGSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800
 Db 1741 REEVGVLFQDRNNLRVKSVSVSGSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800

Qy 1801 QDQNTPRRFTLEGGIAQANFQVASALTDLKKEGLEMKS 1838
 Db 1801 QDQNTPRRFTLEGGIAQANFQVASALTDLKKEGLEMKS 1838

```

RESULT 6
AAU04490
ID AAU04490 standard; protein; 518 AA.
XX
AC AAU04490;
XX
DT 23-OCT-2001 (first entry)
XX
DE Erwinia amylovora disease-specific region E (dspE) mutant protein #1.
XX
KW Disease-specific region E; dspE; insect control; disease resistance;
KW hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic;
KW cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr;
KW petunia; chrysanthemum; carnation; transgenic plant; avirulence locus;
KW pathogenicity; fireblight; mutant; mutein.
XX
OS Erwinia amylovora.
OS Synthetic.
XX
PN US6228644-B1.
XX
PD 08-MAY-2001.
XX
PF 22-JUL-1998; 98US-00120663.
XX
PR 06-AUG-1997; 97US-0055106P.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Bogdanove AJ, Kim JF, Wei Z, Beer SV;
XX
DR WPI; 2001-327491/34.
XX
PT Polynucleotides encoding hypersensitive response eliciting proteins or
PT polypeptides useful for imparting disease resistance to plants, to
XX
PS enhance plant growth, and/or to control insects on plants.
XX
PS Example 9; Col 34; 37pp; English.
XX
CC The sequence represents an Erwinia amylovora disease-specific region E
CC (dspE) mutant polypeptide, which elicits a hypersensitive response in
CC plants. The dspE operon functions as an avirulence (avr) locus and the
CC dspE protein is required for pathogenicity of Erwinia amylovora. DspE is
CC also responsible for causing fireblight in certain plants. The protein
CC and its associated nucleic acid can be applied in a non-infectious form
CC to plants or plant seeds to impart disease resistance, to enhance plant
CC growth, and/or to control insects on plants. Alternatively, plant cells
CC may be transformed with the DNA to form transgenic plants with the same
CC properties. The method can be utilised to treat a wide variety of plants
CC and seeds, including crop plants such as rice, wheat, corn, cabbage,
CC cauliflower, garlic, cucumber, apple, grape, tobacco and sugarcane, and
CC ornamental plants such as Arabidopsis thaliana, petunia, chrysanthemum
XX
SQ Sequence 518 AA;
Query Match 28.8%; Score 2723; DB 4; Length 518;
Best Local Similarity 100.0%; Pred. No. 4.1e-163;
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 GSHHEIKKEPVGSTKATTAHADRVETIAQEDDDSEFQQLHQRLARERENPPQPKLGVA 262
DB 1 GSHHEIKKEPVGSTKATTAHADRVETIAQEDDDSEFQQLHQRLARERENPPQPKLGVA 60
QY 263 TPISARFPQKLTAVAESVLEGTDTTQSPKPSMLKSGAGVTPLATLDKGLQAPDN 322
DB 61 TPISARFPQKLTAVAESVLEGTDTTQSPKPSMLKSGAGVTPLATLDKGLQAPDN 120
QY 323 PPAINTLLKQTLGKDTQHYLAHASSDGSQHLDDNKHGLFDIKSTATSVSLHNSHPGE 382
DB 121 PPAINTLLKQTLGKDTQHYLAHASSDGSQHLDDNKHGLFDIKSTATSVSLHNSHPGE 180

```

```

QY 383 IKGKLAQAGTGSVSDGKSGKISLGGTQSHNKTMLSQPGEAHRSLLTGTIWOHPAGAARP 442
DB 181 IKGKLAQAGTGSVSDGKSGKISLGGTQSHNKTMLSQPGEAHRSLLTGTIWOHPAGAARP 240
QY 443 QGESIRLHDDKIHILHPELGVWQSADKDTSHQSLROADGKLYALKONRTLQNLSDNKSSE 502
DB 241 QGESIRLHDDKIHILHPELGVWQSADKDTSHQSLROADGKLYALKONRTLQNLSDNKSSE 300
QY 503 KLVDKIKSYSDQGOVAILLTPTGRHKMSIMPSLDASPEHSLSLHFADAHQGLLHGK 562
DB 301 KLVDKIKSYSDQGOVAILLTPTGRHKMSIMPSLDASPEHSLSLHFADAHQGLLHGK 360
QY 563 SELEAQSAISHGRLLVADSEGLFSAAIPTKQDGNELXNKAMPQHALDEHFGHDHGISG 622
DB 361 SELEAQSAISHGRLLVADSEGLFSAAIPTKQDGNELXNKAMPQHALDEHFGHDHGISG 420
QY 623 FHDDHGQNLALVKNFRQGHACPLGNHDHGFHGMNLTALVIDNQLGLHHTNPEPHEIL 682
DB 421 FHDDHGQNLALVKNFRQGHACPLGNHDHGFHGMNLTALVIDNQLGLHHTNPEPHEIL 480
QY 683 DMGHLSLALQEGKLYFDQLTKGWTGAESDCKQLKKG 720
DB 481 DMGHLSLALQEGKLYFDQLTKGWTGAESDCKQLKKG 518

```

RESULT 7

AAU04491

ID AAU04491 standard; protein; 507 AA.

XX AC AAU04491;

XX DT 23-OCT-2001 (first entry)

XX DE Erwinia amylovora disease-specific region E (dspE) mutant protein #2.

XX KW Disease-specific region E; dspE; insect control; disease resistance;

XX KW hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic;

XX KW cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr;

XX KW petunia; chrysanthemum; carnation; transgenic plant; avirulence locus;

XX KW pathogenicity; fireblight; mutant; mutein.

XX OS Erwinia amylovora.

XX OS Synthetic.

XX PN US6228644-B1.

XX PD 08-MAY-2001.

XX PF 22-JUL-1998; 98US-00120663.

XX PR 06-AUG-1997; 97US-0055106P.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Bogdanove AJ, Kim JF, Wei Z, Beer SV;

XX DR WPI; 2001-327491/34.

XX PT Polynucleotides encoding hypersensitive response eliciting proteins or

XX polypeptides useful for imparting disease resistance to plants, to

XX enhance plant growth, and/or to control insects on plants.

XX Example 9; Col 34; 37pp; English.

XX CC The sequence represents an Erwinia amylovora disease-specific region E

XX (dspE) mutant polypeptide, which elicits a hypersensitive response in

XX plants. The dspE operon functions as an avirulence (avr) locus and the

XX dspE protein is required for pathogenicity of Erwinia amylovora. DspE is

XX also responsible for causing fireblight in certain plants. The protein

XX and its associated nucleic acid can be applied in a non-infectious form

XX to plants or plant seeds to impart disease resistance, to enhance plant

XX growth, and/or to control insects on plants. Alternatively, plant cells

XX may be transformed with the DNA to form transgenic plants with the same

CC properties. The method can be utilised to treat a wide variety of plants
 CC and seeds, including crop plants such as rice, wheat, corn, cabbage,
 CC cauliflower, garlic, cucumber, apple, grape, tobacco and sugarcane, and
 CC ornamental plants such as Arabidopsis thaliana, petunia, chrysanthemum
 CC and carnation
 XX
 SQ Sequence 507 AA;

Query Match 27.1%; Score 2563; DB 4; Length 507;
 Best Local Similarity 100.0%; Pred. No. 5e-153;
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1064 TVLGHQGVLSKNGEINSEFPSPGKALVQSPNNRSGDLSKSLQQAQVHATPPSAESKL 1123
 Db 1 TVLGHQGVLSKNGEINSEFPSPGKALVQSPNNRSGDLSKSLQQAQVHATPPSAESKL 60
 QY 1124 QSMGLGHFVSAGVDSHQGEIPLGRQDPNDKALTAKSLILDVTTIGELHELADKAKLV 1183
 Db 61 QSMGLGHFVSAGVDSHQGEIPLGRQDPNDKALTAKSLILDVTTIGELHELADKAKLV 120
 QY 1184 SDHKPDADQIKLQROQFDTLREKRYESNPVKHYTDMGFTHNKALEANDAVKAFINAFKK 1243
 Db 121 SDHKPDADQIKLQROQFDTLREKRYESNPVKHYTDMGFTHNKALEANDAVKAFINAFKK 180
 QY 1244 EHHGVNLTTRTVLSQSGAEIAKKLNTLLSDGESMSFSRSGGVSTVFVPTLSKKV 1303
 Db 181 EHHGVNLTTRTVLSQSGAEIAKKLNTLLSDGESMSFSRSGGVSTVFVPTLSKKV 240
 QY 1304 PVPVPIGAGITLDRAYNLSFRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTTSA 1363
 Db 241 PVPVPIGAGITLDRAYNLSFRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTTSA 300
 QY 1364 GNASDWLSAKHKISPDLRIGAAVSTLQGTQLQNSLKFKLTDELPFGFIHGLTHGTLTPAE 1423
 Db 301 GNASDWLSAKHKISPDLRIGAAVSTLQGTQLQNSLKFKLTDELPFGFIHGLTHGTLTPAE 360
 QY 1424 LLQKGIHOMKQSKLTFSDVTSANLDIRAGINLNEDGSKNGVTVARVAGLSASANLAA 1483
 Db 361 LLQKGIHOMKQSKLTFSDVTSANLDIRAGINLNEDGSKNGVTVARVAGLSASANLAA 420
 QY 1484 GSRERSTTSGFGSTTSASNNRPTFLNGVGAGANLTAALGVAAHSTHEGKPVGIFPAPTS 1543
 Db 421 GSRERSTTSGFGSTTSASNNRPTFLNGVGAGANLTAALGVAAHSTHEGKPVGIFPAPTS 480
 QY 1544 TNVSAALALDNRSTQSISLELKRAEPPV 1570
 Db 481 TNVSAALALDNRSTQSISLELKRAEPPV 507

RESULT 8

AAV69258

ID AAV69258 standard; protein; 201 AA.

XX

AC AAV69258;

XX

DT 30-MAY-2000 (first entry)

XX

DE Amino acid sequence of the secretion signal of the DspE protein.

XX

KW Type III secretion system; DspE protein; hairpin secretion system;

KW effector protein; Avr protein; avirulence protein; agriculture.

XX

OS Erwinia amylovora.

XX

FN WO200002996-A2.

XX

PD 20-JAN-2000.

XX

PF 09-JUL-1999; 99WO-US015425.

XX

PR 10-JUL-1998; 98US-0092357P.

XX

PA (CORR) CORNELL RES FOUND INC.

XX

PI

XX

DR

XX

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity

Matches 201;

Conservative

Mismatches

Indels

Gaps

0;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Bauer DW, Beer SV, Bogdanove AJ, Collmer A, Ham JH;

WPI; 2000-182218/16.

New DNA construct encoding type III secretion system, used for

recombinant production of secreted protein and for identifying potential

effector proteins.

Claim 6; Page 14-15; 71pp; English.

The present sequence represents a functional type III secretion system

from the N-terminal domain of the DspE protein. The DspE protein is

secreted by a type III secretion system. The DspE secretion signal is

compatible with the hairpin secretion system of Erwinia amylovora. The

secretion signal polynucleotide is used to produce the constructs of the

invention, which also comprise a promoter and a second DNA that encodes a

protein or polypeptide that can be secreted by the type III secretion

signal. Host cells containing the constructs are used for recombinant

production of the polypeptide. The constructs are also used to screen for

potential effector proteins, e.g. Avr (avirulence) proteins potentially

useful in agriculture

Sequence 201 AA;

Query Match

Best Local Similarity

Matches 201;

Conservative

Mismatches

Indels

Gaps

0;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY	1284	SRSYGGGVSTVFVPTLSKK-----VPVPVPGAGIT-----LDRAINLSFS	1324
Db	1746	IR--GGKVDLTAQQTINRGRIQGDORVTLKGRDITSASTVRGEANRMLDPAGIYVQ	1803
QY	1325	RTSGGLNY-----SFGRDGGVSGNIMVATCHDV-MPYMTGKTSAGNASDM--LS	1371
Db	1804	NDNGTSLNALNNVQLTASDVKNAGKDGHTEITAGHNLDTLSTRTYEQGN---WGKDN	1860
QY	1372	AKHKISPLRIGAAVSGTLQGTQLNSLKPKLTEBELPFGFIHGLT-----HGTLTTPABELQ	1426
Db	1861	TRH-LTQOODTGSQITGAGEVITLQAGQDLNATAAHVANAGQH-LTAQAGNLLTITGTGASS	1918
QY	1427	KGIEH--QMKOGSKLTFESVDTSANLDRAGINLNEDGSKPGVGTARYSAGLSASANLAAG	1484
Db	1919	DLVEHSQTSKGLWSKSSVETHDEVHQRQALSITTFSGDK---VTLAGKDLNIRGNSVAG	1975
QY	1485	SREASTSG-QFGSTTSASNNRPFTL-----NGVGAGANLTAALGVA-----HSSTH	1530
Db	1976	TQDVSLNAGHOLTVTTAESHDETHLRQEKSGLMGTGGIGFTLGKASQKVTTSDSDQLS	2035
QY	1531	EGKPVGIPFAPTSNVSAALDNRTQSISLELKRAEPTVSDNIS-----ELTSTLKG	1584
Db	2036	KGSTVSGSQGVNUNAGE-----QLRVHGVSEVIAGKDLTLIAGQOVDTTSAENR	2083
QY	1585	HFKDSATTKMLAALKELDDAKPAQLHILOHFSAKDVWGDERYEAVRNKLKLVIRQAA	1644
Db	2084	HHTTTKTEQKSGLTVALSGTAGGAVNSAVQTAAAHNESDPVKALONTKAALSGVQAV	2143
QY	1645	DS-----HSMELGSGASHSTTVNNL-----	1663
Db	2144	QAGRLAEAQGSDDKGNNNLAGVLSYGRQSRSGEQHRQTTQOGSHLTAGDNLITAKGD	2203
QY	1664	----SRINNDGIVE-----LLHKHFDAALPAS--SAKRLGEMMNDPAL-----	1701
Db	2204	GKGASGQGIIRIQGSQLQAGKQLQNAHRDIQLSSQNTQETTKGNSSHGSGALGVGLTA	2263
QY	1702	-----KDIIKQIQSTPFGSSASV	1718
Db	2264	GPQGTGLNVANSVSRGNHENGNGVSHNTTLQAGQTVGLNSGRD--TTLKGAQVSGEQI	2321
QY	1719	SMELKQGL-----RQOTEKAILDGKVGREEVGV--LFQDRNNLRVKSVSVSQ	1763
Db	2322	TAEVKRHLTLSSBOOSORYDSKQONASAGVSATVGPLTNGTASLNASRNKLHNSYDSVOE	2381
QY	1764	SVSKSEGFNTPALILG-----TSNSAAMSMEERNIGTINFKYQGDQONTERRFTLE-	1812
Db	2382	QTGLFAGKGGYQVNVGDHTQLDGAIVASQADTKNTLNTGLTFKGIQIQO---ADFTVEQ	2438
QY	1813	--GGIAQANQVVASALTDLKKGL	1834
Db	2439	QSAGVSLGQTAGQVNLNLA VNGL	2462
RESULT 10			
ABM68759			
ID	ABM68759 standard; protein; 2975 AA.		
XX	ABM68759;		
XX	20-NOV-2003 (first entry)		
DT	Photorhabdus luminescens protein sequence #1856.		
DE	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;		
XX	detection; food; gene expression; plant; animal; microorganism; toxin;		
KW	antibiotic; biopesticide; virulence factor; disease model; plague;		
KW	whooping cough.		
XX	Photorhabdus luminescens.		
OS	WO200294867-A2.		
XX			
XX			
XX			

28-NOV-2002.

07-FEB-2002; 2002WO-IB003040.

07-FEB-2001; 2001FR-00001659.

(INSP) INST PASTEUR.
(CNRS) CNRS CENT NAT RECH SCI.

Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
Buchrieser C;

WPI; 2003-148459/14.

Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
useful e.g as therapeutic antimicrobials and agricultural pesticides.

Claim 2; SEQ ID NO 1856; 1205pp; French.

The invention relates to the isolation of genes and their encoded
proteins from Photorhabdus luminescens. The isolated sequences are
sources of probes and primers for detecting the genome of P. luminescens
and related species; to study polymorphisms; for gene analysis and for
detection/amplification of the genes. Antibodies (Ab) raised against the
polypeptides encoded by the genes are used for detection/identification
of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
carry a gene-containing vector are used to select compounds that
modulate, regulate, induce or inhibit expression of the genes in plants,
animals or microorganisms other than P. luminescens and are able to alter
response or sensitivity to toxins and antibiotics produced by P.
luminescens. Cells transformed to express the genes are useful for
recombinant production of the proteins, particularly toxins and
antibacterials useful as insecticides, bactericides and fungicides. The
genes, proteins, vectors containing the genes and Ab are also useful
therapeutically (to treat microbial infection by bacteria or fungi that
are sensitive to P. luminescens-encoded toxins or antibiotics) and as
biopesticides. Other uses of the genes and the proteins are as virulence
factors and for identifying targets of human diseases for which P.
luminescens is a model (particularly plague and whooping cough). This
sequence represents one of the isolated P. luminescens proteins

Sequence 2975 AA;

Query Match 3.1%; Score 295.5; DB 6; Length 2975;
Best Local Similarity 19.5%; Pred. No. 1e-08;
Matches 47; Conservative 293; Mismatches 855; Indels 799; Gaps 116

QY 23 GHGVALQGSSSSPQNAASLAEGKNGRMRIHOPSTAADGISAAHQ--KKSF--- 77
|||: : : : : : : : : : : : : : : : : :
236 GHGM--DSTQQSYTDIARSVAINKLHAQDLKV--TTGRNIVDAHQVEKKSVDDE 289
: : : : : : : : : : : : : : : : : :
78 -----SLRGCLGTKEFSRPAQGPQTTHSKGATRDLLADDGETQHEA---A 123
: : : : : : : : : : : : : : : : : :
Db 290 KHPAFALDVAAALGMYAHK-----IR-LIGTETGVGVHNAGNTGA 328
: : : : : : : : : : : : : : : : : :

QY 124 APDAARLTRSGVKRRNMDMAGRPMVKGGSGEDKVPT-----QQRKHQLNNFQG 173
: : : : : : : : : : : : : : : : : :
Db 329 SAGEFHITAEGRIENRG--TLSSRDITQLTSSADVTNTGKLSSQSAVNLOAKGALNNGR 386
: : : : : : : : : : : : : : : : : :

QY 174 MR----QTMUSKWAHPA-----SANAGDRLOHSDPP---HIPGSHHETKEBPVGSTSMT 221
: : : : : : : : : : : : : : : : : :
Db 387 VEARGDTTVTAGTIHGSHDSVMAAGLDNDGNTTRPGSLTLTAQHQAQKGKGLATNLTA-- 444
: : : : : : : : : : : : : : : : : :

QY 222 AHADRVIAQEDDDSEFQOLHQOQLARERNPQPPLKVATPISAPFPQPKLTA----- 275
: : : : : : : : : : : : : : : : : :
Db 445 IHSRQIDLS-----DSQTAAGIQLTAGQS-----GISTARASVNAORLTKATPQQF 491
: : : : : : : : : : : : : : : : : :

QY 276 -----VAESVLEGTDITQSPLKPQSMKXGSGAVTPL-----AVTLDKGKLQLA 319
: : : : : : : : : : : : : : : : : :
Db 492 NNDGGQVARAI---HLTTPDLSNQOGKINQGTGELTLHRTLNNREGTVFNQGLTLT 548
: : : : : : : : : : : : : : : : : :

QY 320 PNPPALN-----TLLKQTLGKTQHYLAHHASSD-----GSQHLLEDNRGHLEFDIKST 368
: : : : : : : : : : : : : : : : : :

549 TDR---LNNRQGTIASQ--GEDL-HLTAHQADNNQGTQVLQAGNKGKLSLNTQRWLGD-----598
 369 ATSYSVLHNSHPGEIKGLAQAGTGSVS-----VDGKSGKISLGSQTOSHNKTMLSQ 420
 599 -----XGKLLTNGTLTIQAGELQNLHAETQAGQITINADTLSHQSGVMQ 643
 421 PGEAHSRLTGIWQHAPAGAPOG-----ESIRLHDDKIHLHPGLWQWQSAKDTHSOLS 476
 644 WGDLDLSLTLRLDNDHSGTIAAGNGLNLTATVDNRHGNIVAADQGSGLKTLTKDT---LD 700
 477 ROADGKL---YALK-----DNR-----TLQNL-----496
 701 NQSGLEAGHALQLSATQDNRGSIIVAAGDSATLTVGKTIQNAHGHLEAOTRLTTTSQ 759
 497 ---DNKSSEKLVKIKSVS-----VDQGOVAI-----LTDTPGRHKMSIMPSLDA 539
 760 TLDNTQGVLLAQNIDQGTTHPTNTAGQVIAEDTLTVNSGQLDNTAGLLQAGREMAVDT 819
 540 SPESHISLSLHFADAHQGLHKGSELEAQAQVAISH-GRLVVADSEGLFSAAIPOKQDGN 598
 820 --HGHGLTNTTHADQKAGRLSGQLTLRTGDDINTGGMIAADGKTLTSTAL-----N 871
 599 ELKMKAMPQALDEHFHGDHQISGFHDDHGQNLNVLNKNFRQOHACPLGNDHQFHPGWN 658
 872 NTQGTQIAGNGGLDIH-----SQQLINREGTLQSDALTL-----906
 659 LTDALVIDNOLGHLHTNPPEHILD-----MGHL--GSLAL-----Q 693
 907 -TGQLLDNOG-----QIMGEKTTITSGPLDNRHGHLOGGQLAIDTRHAALDNR 956
 694 EKLHVPDQLTKGWTAESCKQ-----LKKGLDGAAYLLKDGVEVKRLN---IN 739
 957 DGKLLSTDTLTNLTHQDNRHGQVAVGNTVLNVKTQDNTGGLRGTLQTLTNTAHLN 1016
 740 QSTSSIKHGE-----NVFSLPHVRNKPPEGDALQGLNKDDKAQAMAVGVNKKYALTE 793
 1017 RETAHTDNGLEAQNLTNAQOVDMTOGALRAANHLQOVSHTELENAOGLISAGKQLTVGD 1076
 794 KGDIRSPOI--KPTQQLERPAQTLREGISGE-----LKDIDHVKQNLXA 838
 1077 ATPASLVINNRQTLTAGQCA-TINAHGLSGDQLLSQGDMAVLTLEDPH--HTGNTAA 1133
 839 ---LT-----HGEVPHQPREAWQNGAE-----SSWHKALPOSESKLSLDM 879
 1134 NGNLTTLKTGNLINDROIKAGQTLHLDAHNLTNLSASGEISAGQTOIQVHDTLNTTGLIDG 1193
 880 SHEKPIATPED-----QSOHOLKAGWH-----AYAAPERGEPLAVGTSGSQ-- 921
 1194 GLTHLTANTLNNMTGRIYGDQLALQTLTNNTAONGKAAVIAARDRLDITGTGLNNSDH 1253
 922 ---TVFNRLMOGVKGVIPGSGLTVKLSAQGTGMAEGRKVSSEKFSERIRAYAFNPTM 977
 1254 AQIYVGDHLHGGRLDNTLTATQARVLNHAATIEAGRNLIQADRINNNTNAGLVTQV 1313
 978 STPR-PINKAAYA---TOHGHQ-----GREGLKPLYEQ-----GA 1009
 1314 ETEKSPHDAVLSGQTTTRYDMSQVDTSRHNKYGVHDAIMPDGSGNDFEYQYTRVRET 1373
 1010 LKOLD-----AHNVRNAPOPDLQSKLETLIDGEGHAELLNDKMRFRDEQSAT--- 1060
 1374 QVKOSDPEKILAGNITLNSAQVTHDS-QIVACGTLGGEI-----GELHNIATQGE 1424
 1061 RSVTVLG-----HQGVLSKNGEINSEFKPSG-----KALVQSFNVNRS 1100
 1425 RITTDGSGTHWYAKKRLKPRFGTKTSQKSRVAPAPVETIDLKTAWQAHTRPQ 1494
 1101 QDLKSLSLOQA--VHATPPSAESKLSQMLGHFVSAGVDMHSHQKEIP-----1145
 1485 GTDITITDROGTQIHAAPTA-----VTFVNGINNQLVLPQGFELSLPPT 1532
 1146 -LGRQRDNDKTLTKSL---ILDVTIG-ELHELAD-----KAKLVSD-----1185
 1533 VKGQTVDPVIRVVPDTRLPNNSLTVQPGSDSHVLVETDPKFTQYKOWLGSDYMRQLT 1592

1186 -----HKPAD---QIKOLROQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKA 1236
 1593 HDPALVHKRLGDFEQRDLVRDQITQITQGVYLS-----GYND---EAQKALMD 1640
 1237 FINAFKKEHI---GVN-----LTRTVLESQGSABELAK-----1266
 1641 AGVAFKQOQLTPGVALSAPAMALLTSDIWLNTQVTLPDGTTEVTVTVQVVARVROGD 1700
 1267 -----KLXN-----TLLSLDSGESMSFSRSYGGVSTVFVPT 1298
 1701 LRSDGALLAGNTVALNNOGDIINSGTISGRDVTQLTANNLTNSGFIR--GGKVDLAAQQT 1758
 1299 LSXK-----VPVPVTPAGAIT-----LDRAVNLFSFTSGGLNVSF-----1334
 1759 LTRNGGQIQDGRDVTILKGRDITSASTVRGDEANRWLDRFAGIYVQNDKGTLSLSAINNVQ 1818
 1335 -----GRDGVGSGNIMVATGHDV-MPYMTGKTKSAGNASDW--LSAKHKISPDRLRGA 1386
 1819 LTASDVKNAGKOGHTEITAGHNLTLDAALSTRTEQG---DWKDNTRH-LTQOODIGSQI 1874
 1387 SGTQGTQNLQSLFKLTDELPGFIHGLTHG-----TLTPAELLQKGIHQMKQSGK---1438
 1875 TGTGEVTLQAGQDLNATAAHVNAGQHLTAQAGNSLTLTCTAGSDILVEHS-KQTSKGWLS 1933
 1439 -----LTFSDVTSANLDRAGINLNEGSKPENG-VARVSAGLSASAN 1480
 1934 KSSVETHDEVDHRAQLSTTFSGD---KVNLAQKDLNIRGNSVAGTQDVSLNAGHQTJVT 1990
 1481 LAAGRSERS-----TTSQ-----QFGSTTSASNRRPTFL 1509
 1991 TAAESDHEHLROBKXGLMGTCGIGFTLCKASQKVTITSDRQLSKGSTVSSQGNVTLN 2050
 1510 NG-----VGAGANLTAALGVAHASTH-EGKPVGIFPAPTSTNVSAAL 1550
 2051 AGQLRVHSGEVIAGKDLTLTGQVDITSANRHHHTTTKTEQKQSGLTVALS---CAAG 2106
 1551 ALDNRTSQISLSEKRAEP---VTSNDISELTSLGKHFKDSATTMLAALKELDD---1603
 2107 GAVNSAVOTARAARTESDPRVKALONTQAALSGV-----QAAQAGRLAAQSGDDKGN 2160
 1604 -----AKPABOLH--ILQO--HPSAKD-----VVGDER-----1627
 2161 NLAVLSYSGRQSRSEQHRQTTQOGSHLTAGDNLITAKGDDKGSQNGDIRIQSG 2220
 1628 YEAVRNLKLVIRQ-QAADSHMELGSASHST-----TYNNLS-----1664
 2221 LQAGKDLQNAHRDILQSSQNTETQTKNSSHSGSALGVGLTVGPGGTGLNISANVSRGN 2280
 1665 -RINNDGIVELLKHFDPAALPASSAKELGEMMNNDPAKDIIKQLOSTPSSASVSMELK 1723
 2281 GRENGGV-----SHTNTTLQA--GQTVGLNSGDRDTLK-----GAQVSEGFQITAEVK 2326
 1724 DGL-----REQTEKAILDGKVGREEVGV--LFQDRNLRVKSYSVSQSVSKS 1768
 2327 RHLTSLSEQDSQVDSQQONASAGVSATVGLTNGTASLNASRNKLSHNVDSVQEQTGLF 2386
 1769 EGFNTPALILG-----TNSAAMSWERNITGTINFKYQDQNTPRFTLE---GGI 1815
 2387 AGQGGYQVNVGDHFTQLDGAVIASQADKANNTLTNTGLGFK---DIKNQADFTVBEQSAGI 2443
 1816 AQANPQVASALTDLKKEGL 1834
 2444 SLQGTAGVLLNLAUVNGL 2462

RESULT 11
 ABM67717
 ID ABM67717 standard; protein; 2951 AA.
 XX
 AC
 XX ABM67717;
 DT 20-NOV-2003 (first entry)

QY	1114	ATPPAESKQLQSMGLHFVSAGVDMSHQKEIP-----LGRQDPNDKTALTAKSRL-----ILD	1166
Db	1370	SAFTA-----VTPVNGIKNQPLVLPETVKGTVDVIRVVTPDRLPNNSLY	1417
QY	1167	TVTIGELBELADAKALVSHKPDADQIKQ-----LRQPF-----DTLEKRY	1208
Db	1418	TVQPG-----SDSHYLV-ETDPKFTQYKQWLGSDYMRQQLTHDPALVHKRLGDFGFEQRL	1471
QY	1209	ESNPVKHYTDM-----GFTINKALEANDYAVKAFINAFKKEHH-----GVN	1249
Db	1472	VRDQITLGTQORVLPQYNNND-----EAQFKALMDAGVAFQKQQLTPGVALSPAQMALLTSD	1528
QY	1250	---LTPRTVLESQGSALA-----KKLXNTLLSLD-----S	1277
Db	1529	IWLNTQVTLPGTTEVVTPQVYARVRQGDLSGALLAGNTVALNSQGDITNSGTMS	1588
QY	1278	GESMS-----FSRSYGGGVSTVFPILSKK-----VPPVFIQAGIT-----	1314
Db	1589	GRDVLTQLTANNLTNSGFIR--GGKVDLAAQQTLTNRGGQIQGDDRVTLKGRDITSASTR	1646
QY	1315	-----LDRAYNLFSRSTGGILNVGF-----GRDGGVSGNIMVATGHDV-MPYM	1356
Db	1647	GDENRWLDPRAGIYQNDKGTLSLSAINNVQLTASDVKNAGDKGHTETITAGHNLTLDAL	1706
QY	1357	TGKKTSGNASDW--LSAKHKISPLDLRIGAAVSGTLOQTQNSLKFKLTEDEL-----	1407
Db	1707	STERTEQG--DWGKDNTRLTQQQDQIGSQITGAGEVTLQAGQDLNATAAHVNAVQQLTA	1763
QY	1408	PGFIHGLTHTLTPAELLQKGTIEH--QMKQSKLTFSDVTSANLDRAGINLNEGSKP	1464
Db	1764	QAGNLTLTATGTAS--SDL-----VEHSKQTSKGLWSKSSVETHDEVDRQALSTTFSGDK-	1817
QY	1465	NGVTARVSAGLSASANLAAGSRSTTSG-QFGSTTSASNRRTEFL-----NGVGAGANL	1518
Db	1818	---VTLQAGKDLNIRGNSVAGTQDVSLNAGHQLTVTTAAEAHDETHLRQEKKSGLMTSGM	1875
QY	1519	TAALGVA-----HSSSTHEGPVGIFFAFTSTNVSAALADNRSTSQSISLBELKRAFPV	1570
Db	1876	GFTVGRKASQVTTSDSDLSQSKGSTVSGSGNVTLNAGE-----QLRVHGVSEVI	1923
QY	1571	TSNDIS-----ELTSLTGLGHFKDSSATTMLAALKELDDAKPABQLHLOOHFSKADVVG	1624
Db	1924	AGKDLTLTGQOQVDTLSAENRHHHTTKTEQKQSLTVALSAGAGAVNSAVQTAARTES	1983
QY	1625	DERVEAVRNKLKVIROQAADSHSMELGSHASTVYNLNRIN-----NDGIVELLHK--	1677
Db	1984	DPRVKALONTKAALSQVQAAGLAEAGQSDDKGNNNLAGVLSYGRQSRSEQQHROT	2043
QY	1678	-----HFDA-----ALPASSAKRLGEMMNNDPALKNDIIKQLQSTFFSSASVSMELKGLRE	1728
Db	2044	TQOGSHLTAGDNLITITAKGDKGASGQNGDIRIQG--SQLQAGKDLQLNANRDITLSSSQ	2101
QY	1729	QTEKAILDGKVGREEVGLFQDRNRLAVKSVSVSQSKSEGFNTPALLLGTSNSAAMS	1788
Db	2102	NTEQT-----TGKNSHRSGLMGLTVGPG-----GTGLNVSANV	2136
QY	1789	ERNIGTIN-----FKYQDQ--QNTPRRFTLEGGIAQANPQVASA/TDLKK	1831
Db	2137	SRNGRENGVSHVHTNTLQAGQSVGLNSGRDITLKG--AQVSEQITA--DVKR	2187
RESULT 12			
ID	ABM67454		
XX	ABM67454	standard; protein; 4560 AA.	
AC	ABM67454;		
XX	20-NOV-2003	(first entry)	
DT	DT	Photorhabdus luminescens protein sequence #551.	
DE	DE	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;	
XX	XX		

detection; food; gene expression; plant; animal; microorganism; toxin; antibiotic; biopesticide; virulence factor; disease model; plague; whooping cough.

Photorhabdus luminescens.

WO200294867-A2.

28-NOV-2002.

07-FEB-2002; 2002WO-IB003040.

07-FEB-2001; 2001FR-00001659.

(INSP) INST PASTEUR.

(CNRS) CNRS CENT NAT RECH SCI.

Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A; Buchrieser C;

WPI; 2003-148459/14.

Genomic sequence of Photorhabdus luminescens and encoded polypeptides, useful e.g. as therapeutic antimicrobials and agricultural pesticides.

Claim 2; SEQ ID NO 551; 1205pp; French.

The invention relates to the isolation of genes and their encoded proteins from Photorhabdus luminescens. The isolated sequences are sources of probes and primers for detecting the genome of P. luminescens and related species; to study polymorphisms; for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than P. luminescens and are able to alter response or sensitivity to toxins and antibiotics produced by P. luminescens. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and antibacterials useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically (to treat microbial infection by bacteria or fungi that are sensitive to P. luminescens-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which P. luminescens is a model (particularly plague and whooping cough). This sequence represents one of the isolated P. luminescens proteins

Sequence 4560 AA;

Query Match 3.0%; Score 280.5; DB 6; Length 4560;

Best Local Similarity 19.9%; Pred. No. 1.7e-07;

Matches 440; Conservative 247; Mismatches 786; Indels 735; Gaps 113;

QY 29 QCGSSSSPQNAASLAACEKNEKMPRIHQPSTAADGISAAHQKSKFSLRGCLGPKKF 88

Db 1813 QCGRVITADKOLNVTSTGAVDNTSGKIVSQHQQLTMNTG-----ELNNTSGLLQSKTT 1864

QY 89 SRGAPOQPGQTHSKGATLRDLARDGGETQHEAAAPD--AARLTRSGGVK--RRNMDDM 144

Db 1865 LYLNTHGQKLTNTQSG---DDLGRSDSLTLEAGEIDNTAGKIDSQGETKLTSLNNT 1921

QY 145 AGRPMVKG-----GSGEDKVFTQQR----- 165

Db 1922 DGKILSKGKADLTQAINNORGLTQSASLKLDTQQQKLTNTDTSGLNGLALDLSKLIT 1981

QY 166 -HOLNFEQMRQTWLSKWAHPASANAGRLQHSPPHIPGGSHETKEKPPVGSKTATTAHA 224

Db 1982 GELINDFGYIR-----GDE-----THI-NSHQQTLSNLTGTIVSKKNLKL 2020

QY 225 DRVEIAQ-----EDDDSEFOQLHQORLAREFENFPQPKLGVATPISARFQPKLTAVAES 279

```

Db 2021 DSGELDTGGLIKSDGNMTIDTHGKKLTAKSGDTR-----GVIS-----EGTMTLTADDE 2070
Qy 280 V-----LEGDTTQSPKQPMKSGAGVTPLAWTLDKGLQPLADNPPALNTLLKOT 333
Db 2071 IDNQDFIKGTGT---IVTGGELKNQGG-----TLISEKSLTISVQNTDSGGLL-QS 2121
Qy 334 LQK---DTQ-HYLAHASSD-----SQHULLDNKKGHLFDIKSTATSYSV-----LHNSHP 380
Db 2122 AGKLTLDTHGSLTNKNSGDRGIRSQDDMLITSGDLHNOAGTITNRKIATVNNLQNTIV 2181
Qy 381 G-EIK-----GKLAQAG-----TGS----- 394
Db 2182 GTQQLKLTQAPNNQGGVHSDGNLNLNTQGRNLANTGDKSGSFSAQGDLTLDIGELN 2241
Qy 395 -----VSDGKS-----GKISLGSQTSHNKTMLSPG-----EAHRSI 428
Db 2242 NDASFIADGKTITSTLTNNKGLIAGNSGLIEHISQTLTNSGSLKSNANTVINDTNGQL 2301
Qy 429 LTGIWQHAPAGARPAGESIRLHDDKIHILHPGLVWQSAKDTHSGLSRQADGKLYALKD 488
Db 2302 LDNQHGRIIGDNTTWTBGLNNQHGHIOGKKLTI-NTQADTDNQ-----DGKLLS-AD 2354
Qy 489 NRTLQNL- DNKSSEKLVDKISYVDORGVAIILTD-----TPGRHKMSIMPS 536
Db 2355 TNNLTLOLDNR-----RGVKAIGDATINAPKONTNNTGGLIHSDQQLT 2398
Qy 537 L-----DASPESH-ISLSLHPADAHQGLLHGKSELEA---QSWAISHG----- 575
Db 2399 LKTABLINRETNHPDQGTAEADLIEAQIDNTQGTLOGANPLQALINQSLKNEQGLISG 2458
Qy 576 --RLVADSEKLSAATPKQ--DGNE-LKMKAMPQHALDEHFGHDHQISGFFPHDDHQ 630
Db 2459 GKQTIETDKTQKL--TVNNQGTITGSEKVIKA---NAL-----SGDGQIL-----SQGD 2504
Qy 631 LNALVKNFRQQAACPLNGDHFHFGWNLTDALVTDNQLGLHHTNPEHEILDMGHLSL 690
Db 2505 IEVKLKQDFHN-----TGNIAADGKLSLETD-----GNIIND 2536
Qy 691 ALQEGKLYFD--QLTGWTGAESCKQ---LKKGLDGAAYLLKDGVEKRLINQSTSS 744
Db 2537 STMKAEQAYLEAQNLTNTQT-AEIRAKOTEVNVNRLTNTG--LIDGELTHLTANKVLDN 2593
Qy 745 IKHGTENVFSLPHVRNKPEPGD-----ALQGLNKDKQAQAMAVIGNKYALATEKGD 797
Db 2594 T--GTGRYI-----GDQIALKSGTLNNTAKDGKA---AVIAARDRL----- 2629
Qy 798 RSFOIKPQTQQLERPAQ--TLISREGISGELKDIHVHDKONLYALTHEGEVHPQPREAQ 855
Db 2630 ---DIGTGLNRRHQAQIYSVDMRIGQGL-----DENLAATSQAGRL----- 2669
Qy 856 GAESSWHKLALPOSES-KLXSLDMSHEHKPIAT----FEDCSQHLKAGG-----WHA 904
Db 2670 -----SNEAATIEAGSLKIDAAIINKNTNGLVTDIVETEKSORHEAVLSGRTARYDWSQ 2724
Qy 905 YAAERGPPLAY-----GTSSQTVF-----NRLMQGVKGVIPGSGLTVKLSAQ 948
Db 2725 VNTSRNKYGVHRAVMPDGSSEEFYRYRTTINETQIKESDPGKILSGNNTINSAML 2784
Qy 949 T-----CGMTGAE-----GRKVSXKFSERIRAYA-----FNPTMST----- 979
Db 2785 TNDISQIVAGILGDDIDELHNNATKGERIITDEGSEIRWYAKKKRRFRGKTSSQGDW 2844
Qy 980 ----PRPIKNAAYATQHCWQREGKLPLEYMQGALIKQLDADHNVNRHNAQPQDLOSKLETL 1035
Db 2845 DHYNPATITETDLKALAWQG--NTRP--NSTGITIDDRQTSRVQSTPTGILNLSRM--A 2898
Qy 1036 DLGE-----HGAELNDMKRFRDELEQSATRSVTVLQGHQGVKNSGEBINSEFKPSP 1087
Db 2899 EAGEVFTFNATINAILPN-----RDIPTDRPLLSPT--GQOTEQLLSSGAV-----ALP 2946
Qy 1088 GK-ALVQSFNVNRSGD-----LSKS-----LQQAQVHATPPAESKIQ 1124
Db 2947 SRDALTEHPTLUSPTGQQTQEVQLSSSAVVLPNRDALTEHPVLPVPOQTTERVLPP----- 2999

```

RESULT 13

AA75097

ID AA75097 standard; protein; 2514 AA.

XX AC AA75097;

XX XX AA75097;

XX DT 21-MAR-2000 (first entry)

XX DE Neisseria meningitidis ORF 564 protein sequence SEQ ID NO:1668.

XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

XX KW antibacterial; gene therapy.

XX OS Neisseria meningitidis.

OS

XX WO9957280-A2.
 XX 11-NOV-1999.
 XX 30-APR-1999; 99WO-US009346.
 XX 01-MAY-1998; 98US-0083758P.
 XX 31-JUL-1998; 98US-0094869P.
 XX 02-SEP-1998; 98US-0098994P.
 XX 03-SEP-1998; 98US-0099062P.
 XX 09-OCT-1998; 98US-0103749P.
 XX 09-OCT-1998; 98US-0103794P.
 XX 09-OCT-1998; 98US-0103796P.
 XX 25-FEB-1999; 99US-0121528P.
 XX (CHIRON) CHIRON CORP.
 XX (GENO-) INST GENOMIC RES.
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 XX Tettelin H, Venter JC;
 XX WPI; 2000-062150/05.
 XX N-PSDB; AAZ53859.
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 XX vaccines and diagnostics.
 XX Claim 2; Page 854-855; 1453pp; English.
 XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941
 XX represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 XX and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
 XX PCR primers used in the exemplification of the present invention. The
 XX polypeptides, the polynucleotides, antibodies and compositions of the
 XX invention can be used as vaccines, as diagnostic reagents, and as
 XX immunogenic compositions. The polypeptides can be used in the manufacture
 XX of medicaments for treating or preventing infection due to *Neisseria*
 XX bacteria (e.g. meningitis and septicemia), to detect the presence of
 XX *Neisseria* bacteria, or to raise antibodies. They may also be used to
 XX screen for agonists or antagonists, which may themselves have use as
 XX antibacterial agents. The polynucleotides of the invention may also be
 XX used in gene therapy protocols
 XX Sequence 2514 AA;
 XX
 XX Query Match 2.9%; Score 275; DB 3; Length 2514;
 XX Best Local Similarity 18.8%; Pred. No. 1.5e-07;
 XX Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;
 XX
 XX 7 GTEH-KAAVHTAAHNPV-----GHGVALQQGSSS-----SSPQNAAS 43
 XX 37 GSAHVKSVPFTTAPVCRNIFSFLLGSLCLAVGTANTAFADGIADKAAKPTQAT 96
 XX 44 LAAGCKNRGKMPRIHQPTAAGDISAAHQKKSFSLRGCLGTRKFSRQPGQPG----- 98
 XX 97 ILQTGNG---IPQVNIQPTGAGVSNQYQAFDVGNGAILNN--SRSTNTQTQLGGMIQ 151
 XX 99 -----TTHSK-----GATRLDLLADDGTOHEAAPAAALTR 132
 XX 152 NPWLARGEARVVVQINSSHSQMNQYIEVGRBAEVVIANPAGIANGGGFINASRATL 211
 XX 133 SGGVKRRNMDMAGRPVKG-----GSGED-----KVPTQKRHLNNGFQMRQTMLS 180
 XX 212 TTGQPOYQAGLSGFKRQGNVIVAGHGLDARDTDFRLLSYHSHKIDAPVWGQVVRVAG 271
 XX 181 KWAHPASAGDRLOHSPPHIPGSHHEIKERFVGSTSKATTAHADRVEIAQEDDDSEFQ 240
 XX 272 QNDVVATGNA-----HSP-----ILNNAANTSNNTANGTHIPLFAIDTG----- 312
 XX 241 LHOQLARERPNFPQPLGVATPISARPOKLTAVASVLEGTDTTQSPKPKQSMKGS 300

Db 313 -----KLG-----GMVANKITLISTA-----EQAGIRNOGQJFAS 342
 Qy 301 GAGVTPPLAVTLDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHAS--SDGSQHLILDN 358
 Db 343 SGN-----VAIDANGRLVNSGTMAAN-----AKDTNTAEKVNIRSQG-----VEN 385
 Qy 359 KG-----HLFDIKSTATSYS-----VLHNSHGEIIGKLAOAGTGSVDGKSGKI 404
 Db 386 SGTAVSQOQGTQIHSQSIGTGTLLSSGEILHNS-----GSLKNETSGTI-----EAARL 435
 Qy 405 SLGSGTQSHNKTMWSQGEAHRSLLTGICWHPAGAPQGESIRLHDDKTHILHPGLGV 464
 Db 436 AIDTDT-LNNQGGKLSOTG-----SQUHLI----- 458
 Qy 465 QSADKOTHSQSRQADGKLYALKDNRTQLNSD-----NKSSEKLVDKIKSYSDVDRQG 518
 Db 459 -----DAQGK-----DNRGMLQDTAPTASNGSSNQTN-----SYNASPHSS 498
 Qy 519 VAILTPTGPRHKSIMPSLDASPEHSISLGHFADAHQGLLHGKSELEASVALSHGRVL 578
 Db 499 TTTPTTATGTGTATVSISNITAPT-----FAD-----GTIRTHGALDNGSGSIANGQTD 547
 Qy 579 VADSEKLFSAAIPOKODGNELKMKAMPOHALDEHFGHDHQISGFHDDHGHQNALVKN- 637
 Db 548 VSAQOG-LNNAG---QIDIHQLNAG---SAFDNHNG---TIISDAVHIQAGSLNNQNGNI 598
 Qy 638 NFRQOQHACPLGNDHQFHPGWNLTALVID-NQLGLHHTNPEPEHILDMHGLSLALQEGK 696
 Db 599 TTRQQ-----LEIETDQLDNAHGKLLSAEIALDVLAVSGSLNNQNGE 638
 Qy 697 LHVFDQLTKWTGAESDCKQLKGLDGAAYLLKDGVEKRLNINQSTSSIKHGHTENVSILP 756
 Db 639 IATNQQL-----IHHGQOSTAVIDNTNNTIQSGRDVAIOAK 675
 Qy 757 HVNKEPPEGALQGLNK-----DDKAQAMAVIGNVKVALTEKGDIRSFOIKPGTQOLE 810
 Db 676 SLNS-----NGTLAANDKLDIALQODFVERNIIVAGNE-LSLSRGSLSKNSHTLQAGKRIR 730
 Qy 811 RPAQTLREG-----ISGELKDIHVHDKQNLALTHEGEVHPHPREAWQNGAESSSWHKLA 866
 Db 731 IKANNLDNAOQNIQSGTTDITGQHN-----LTNRGLIDGQ----- 767
 Qy 867 LPQSESKLSLDNSH-----EHPKFIATFEDGSOHLKAGGWHAAAPERGPLAYGTS 918
 Db 768 -----QTKIQAGQMNIGTGRIYGNIAATAALDNDQDENGTTGA--ATAARENLMIGI-- 819
 Qy 919 GSQTVFNRLMQGVKVI PGSLTVKLSAOTGGMTGAEGRKVKSSKFSERIRAYAFNPMS 978
 Db 820 -----QLNNRENSLIYSGNDMAVGGLDNGATGATKAQR----- 853
 Qy 979 TPRPIKNAAYATQHGMOGREGKLPYEMOGALIKQL-----DAHNVRRNAP 1024
 Db 854 -----IHNAGATIEAAGKMWLGVEKHLNTHNBLKTOLVETGREHIVDVEAFGRHELLREGT 909
 Qy 1025 QPDL-----QSKLETLDLGEH-----GALLNDMK 1049
 Db 910 QHELGSVYNDESDEHLETPDGAHENWHKYDEKVTQKTQVOTAPAKIISGNDLTIDGK 969
 Qy 1050 R-REDELEQATRSVTVLGOHQ-----VLKNSGEINSEFKPS-PGKALVQSFN 1096
 Db 970 EVFNTDSQITAGGNLIVQTEKDLHNGQTGEEKKVFSENGKLHSYWEKHKGR--DSTG 1026
 Qy 1097 VNRSQDLSKLSQAOAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGIPLGRQDRPDNDKT 1156
 Db 1027 HSEQNYTLPEITRNI SLGSFAVESHRKALSHHAPSQGTLPQSGN---ISLPTSNSFT 1083
 Qy 1157 ALTKSRL-ILDTVTIGELHELADKAK-----LVSDHKPDA---DQIKQLRQQTDLREKR 1207
 Db 1084 PLPSSSYIINPVNKGYLVTDPFANYROWLGSYMLDSLUKLDNNLHKLGGLGYEQR 1143
 Qy 1208 YESNPVXHYTDMGTGTHNKALEFANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAAELAK 1267
 Db 1144 LINEQIAELT--GHRRLDGYQNDDEEQKALMDNGATAARSNNLSVGIAL-----SAEQVAQ 1197

Db 313 -----KLG-----GMVANKITLISTA-----EQAGIRNQGLFAS 342
QY 301 GAGVTEPLAVTLKQKQLQAPDNPALNTLLKQTLGKDTQHYLAHAS--SDGSHLLLDN 358
Db 343 SGN-----VAIDANGRLVNSGTAAAN-----AKDNTAEKHVNSRQ-----VEN 385
QY 359 KG-----HLFDIKSTATSYS-----VLHNSHPEIEIKGLAQAGTGSVSDGSGKI 404
Db 386 SGTAVSQOQTQIHSQSIGTNTGTLSSGEILLHNS-----GSLKNETSGTI-----EAARL 435
QY 405 SLGSGTQSHNKTMISQGEAHRSLTIGIWOHPAGARPOGESIRLHDDKLIHILPELGW 464
Db 436 AIDTDT--LNNQGLSOTG-----SOKLHI-----458
QY 465 QSAKDQTHSLSRQADCKLYALKDNRTLQMLSD-----NKSSEKLVDKIKSYSDVQRQ 518
Db 459 -----DAQGM-----DNRGEMGLQDTAPTASGSSNOIGN-----SYNASFHS 498
QY 519 VAILTTPGRHKMSIMPSLDASPEHISLSLHPADAHQGLLHKGSELEASQVAISHGRV 578
Db 499 TTTPTTATGTGTATVSIISNAPT-----FAD-----GTIRTHGALDNGSIIANGQTD 547
QY 579 VADSEKLSAALPKQDGNELKWKAMPOHALDEHFECHDQIISGFPHDDHGOALNALVN- 637
Db 548 VSAQCG--LNNAG--QIDIHQLNAG--SAFONHNG--TIIISDAVHIQAGSLNNQNGNI 598
QY 638 NFRQOQACPLGNDHQFHPGWNLTALVID--NOLGLHHTNPEPEIILDMHGLSLAQEGK 696
Db 599 TTRFQ-----LEIETDQLDHAHGLLSAEIADLAVSGSLNNQNGE 638
QY 697 LHYFDQLTKGWTGAESDCKOLKGLDGAAYLLKDGVEKRLINQSTSSIKHGTVENFSLP 756
Db 639 IATNQUL-----IIHDGQOSTAVIDNTNGTIQSGRDVAIQAK 675
QY 757 HVNKEPEPDALQGLNK-----DDKAQAMAVIGVNVKYLALTEKGDIRSFIQKPGTQOLE 810
Db 676 SLSN-----NGTLAADNKLDIALQDDFYVERNIVAGNE--LSLSTRGSLNHSHTIQAQKRIR 730
QY 811 RPAQTLSREG-----ISGELKDIFVHDHKQNLALYTHEGEVPHQPREAWQNGAESSSWHKLA 866
Db 731 IKANNLDNAAQNIQSGGTTDITQHN-----LTNRGLIDGQ-----767
QY 867 LPOSEKLSKLDMSH-----EHPKIATFEGSQHQLKAGGWHAVAAPERPLAVGTS 918
Db 768 -----QTKIOAGOMNIGTGTIRYDNTAIAATRLDQNDENGTA--AIAARENILGTV-- 819
QY 919 GSQTVENRLMOGVKVPISGLTVKLSAQTGCTGAEGKVKSSKTSERIRAVAFNPTMS 978
Db 820 -----QLNNRENSLIYSGNDMAVGALDINGQATGKAOR-----853
QY 979 TPRPIKNAAYATQHGQREGRLKPLXEMOGALIKQL-----DAHNVHRNAP 1024
Db 854 -----IHNAGATIEAAGKMRGLVGEKHLNHTNEHLATQLVETGREHIVDYEAFGREHLLREGT 909
QY 1025 QPDL-----OSKLETLLDGEH-----GAELNDMK 1049
Db 910 QHELGSVYNDESDDLTPDGAARENHWHKYDEKVTQKTQVQTAPAKIISGNDLTDIGK 969
QY 1050 R-FRDELEQASRSTVTLVGHQ-----VLKSNCEINSEFKPS--PGKALVQSFN 1096
Db 970 EVENTSDQIIAGNLIVQTEKDLHNEQTEGKVPSENGKLHSYWEKHGR-----DSTG 1026
QY 1097 VNRSQDLSKSLQQAHAATPPSAESKIQSMLGHFVSAGVDMSHQKQIPLGRQDPNDKT 1156
Db 1027 HSEQNTPLPEITRNISLGSFAYESHKALSHAPSQGTLPQSNQ---ISLPVTSNFT 1083
QY 1157 ALTKSRL--ILDVTIGELHELADAK-----LYSDHKPDA---DQIKQLRQPDPTREXR 1207
Db 1084 PLPSSSLYIINPVNKGVLVETDPRFANYROWLGSYMLDSLKLDPNLHKLRLGDBGYEQR 1143
QY 1208 YESNPVKHYTDMGFTHNKALEANYDAVAFINAFKKEHGHVNLTRTVLESQGSAAELAKK 1267
Db 1144 LINEQIAELT--GHRRLDGVQNDDEQFKALMDNGATAARSNNLSVGIAL-----SAEQVAQ 1197

QY 1268 LKNTLLSLDSGESMSFRSYGGVSTVFVFTLSKKVPPVPIPGAGITLDRAYNLSFRSTS 1327
Db 1198 LTSDIVMLVQKE-----VKLPDGGTQIVLPQVVRVKNKGIDGKGAL-----LSSGNTQ 1247
QY 1328 GGLNVSGFRDGGVSGNIMVATGHDVMPYMGK--KTSA-----GNASDWLSAKHKIS 1377
Db 1248 INVSGSLKNSGTIAGRNALIINTDLNIGRTHAOKSAVTATQDINNIGMLSAEQTL 1307
QY 1378 PDLRIGAAV--SGTLOQTQNSLAKFLTEDELPG--FIHGLTHGTLT-----PAELLQKIE 1430
Db 1308 --LNAGNNINSQSTTASSQNTQSSYVLDRMAGIYITGKEKGLAAQAGKDINIAGQIS 1365
QY 1431 HQMQQSKLTFSDVTSANLTL-----RAGINLINEDGSKPNGVTVARYSAGLSASA----- 1479
Db 1366 NQSEOG--QTFLOAGRDINDLTQVTSKHQATHFDADNHVIRGSTNEVGSSITQKGDVTL 1424
QY 1480 --NLAGSRETSITQSGSTTSASNRPFTFLNGVAGANLTAALGVVAHSSTHEGKPVGI 1537
Db 1425 GNNLNAAAEVSSANGTL--AVSARN-----IN--ISAGINTT--HVDDASKHTGRSGG- 1473
QY 1538 FPAPTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTL-----GKHF 1587
Db 1474 ---GNKLVTIDKAAOSHETHAQSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVR 1530
QY 1588 DSATTTMLAALKELDDAKPAEQJLHILQOHFSAKDVVVDERYEAVRNLLKLVIRQQAADSH 1647
Db 1531 IGT-----QTSQSETVHOTQKSLMSAGIG-----FTI--GSKTN 1565
QY 1648 SMELGSASHSTVYNNLSRINDGIVELLHKKHFDPAALPASAKRLGEMM-----NNDPALK 1702
Db 1566 TQENQSQSHEGTSTGVSLKGDTTI--VAGKHYE-----QIGSTVSSPEGNNTIYAQ 1615
QY 1703 DIILQLOSTPFSSASVSMELKQGLREQTEKAILDKGVGREVG-----LFQDRNLRV 1756
Db 1616 SIDIQAHNKLNSNTTQVEKGLTVAFSSPVD--LAQAIIVAAQSSKQVQSKND--RV 1672
QY 1757 KSVVSQS--VSKSEGTNPALLGTSNSAAMSERNICTINPKYQDQNTPRFT--LEG 1813
Db 1673 NAMAANAAGWQAYQTKSAQNLANGTTNAQVS-----ISITYGEQQN--RQTTQVQA 1723
QY 1814 GTAQANPQVASALTDLKKEGLEMK 1838
Db 1724 NQAQASQIQAGGKTTLIATGAEEQS 1748

RESULT 15

ABU38197
ID ABU38197 standard; protein; 2514 AA.

XX ABU38197;

XX AC

XX XX

DT 19-JUN-2003 (first entry)

XX XX

DE Protein encoded by Prokaryotic essential gene #23724.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Neisseria meningitidis.

XX OS

XX XX

PN WC200277183-A2.

XX XX

PD 03-OCT-2002.

XX XX

PF 21-MAR-2002; 2002WO-US009107.

XX XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:15:47 ; Search time 35.3283 Seconds
(without alignments)
2685.909 Million cell updates/sec

Title: US-09-596-784-2
Perfect score: 9448
Sequence: 1 MELKSLGTEHKAHVTAHN.....NPQVASALTDLKKEGLEMK 1838

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/ECTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9448	100.0	1838	3	US-09-120-663-2
2	9448	100.0	1838	4	US-09-431-614-8
3	1041	11.0	201	4	US-09-350-852A-1
4	257.5	2.7	10182	4	US-09-134-001C-3159
5	238.5	2.5	2504	4	US-09-328-352-5821
6	238	2.5	1589	4	US-09-543-681A-4998
7	236	2.5	3241	4	US-09-841-786-1
8	233	2.5	3696	4	US-09-134-001C-5080
9	216	2.3	2048	4	US-09-268-347-48
10	211.5	2.2	2680	4	US-09-489-039A-7973
11	209.5	2.2	1912	1	US-08-409-995-4
12	209.5	2.2	1912	3	US-08-685-467-4
13	207.5	2.2	2123	3	US-08-540-236-3459
14	205.5	2.2	2123	3	US-08-968-685A-10
15	204.5	2.2	1992	4	US-08-621-944A-3
16	204.5	2.2	1992	4	US-08-945-567D-3
17	203.5	2.2	1780	1	US-08-769-309A-5
18	203.5	2.2	1780	3	US-08-994-570-5
19	201	2.1	2137	4	US-09-134-001C-4463
20	201	2.1	2353	3	US-09-377-155-33
21	201	2.1	2353	3	US-08-913-942-4
22	201	2.1	2353	4	US-09-669-974-33
23	201	2.1	2353	4	US-09-797-862-33
24	197	2.1	2291	4	US-09-252-991A-21854
25	196	2.1	2354	4	US-09-268-347-47
26	195	2.1	2042	4	US-09-077-098A-6
27	195	2.1	2568	4	US-09-866-108A-3

28 194 2.1 2954 4 US-09-150-867-1 Sequence 1, Appli
29 193.5 2.0 2314 4 US-09-268-347-49 Sequence 49, Appli
30 190.5 2.0 1536 1 US-08-038-682-2 Sequence 2, Appli
31 190.5 2.0 1536 1 US-08-302-832-2 Sequence 2, Appli
32 190.5 2.0 1536 2 US-08-530-198-2 Sequence 2, Appli
33 190.5 2.0 1536 2 US-08-469-880-2 Sequence 2, Appli
34 190.5 2.0 1536 2 US-08-728-470-2 Sequence 2, Appli
35 190.5 2.0 1536 2 US-08-617-697-2 Sequence 2, Appli
36 190.5 2.0 1536 3 US-08-719-641-2 Sequence 2, Appli
37 190 2.0 1315 4 US-09-252-991A-22746 Sequence 22746, A
38 190 2.0 2411 4 US-09-268-347-36 Sequence 36, Appli
39 188.5 2.0 1420 4 US-09-125-635-4 Sequence 4, Appli
40 188.5 2.0 1536 4 US-09-206-942-67 Sequence 67, Appli
41 188.5 2.0 2039 4 US-09-077-098A-7 Sequence 7, Appli
42 187 2.0 1600 2 US-08-617-697-10 Sequence 10, Appli
43 186.5 2.0 2482 1 US-08-328-254-6 Sequence 6, Appli
44 185.5 2.0 1761 4 US-09-489-039A-11234 Sequence 11234, A
45 184.5 2.0 1833 4 US-08-621-944A-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-120-663-2
; Sequence 2, Application US/09120663
; Patent No. 6228644
; GENERAL INFORMATION:
; APPLICANT: Bogdanove, Adam J.
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,663
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,105
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1661
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1600
; TELEFAX: (716) 263-1304
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1838 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-120-663-2

Query Match 100.0%; Score 9448; DB 3; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHGVALQOGSSSSPQNAASAAEAGKNGKMPRIHOP 60
 DB 1 MELKSLGTEHKAHVHTAAHNPVGHGVALQOGSSSSPQNAASAAEAGKNGKMPRIHOP 60
 QY 61 STAADGISAHQOQKSFSLRGCLGTGKFFRSAPQOGPGTTHSGATLRLDILLARDGDETOH 120
 DB 61 STAADGISAHQOQKSFSLRGCLGTGKFFRSAPQOGPGTTHSGATLRLDILLARDGDETOH 120
 QY 121 EAAAPDAARLTRGGVKRRNMDMAGBPMVKGSSEDKVPQOQKRLNLFQGMQRTMLS 180
 DB 121 EAAAPDAARLTRGGVKRRNMDMAGBPMVKGSSEDKVPQOQKRLNLFQGMQRTMLS 180
 QY 181 KMAHPASANAGDLQSHPPHPCSHHEIKKEEYVGSSTKATTAHADRVETIAQEDDDSEFQ 240
 DB 181 KMAHPASANAGDLQSHPPHPCSHHEIKKEEYVGSSTKATTAHADRVETIAQEDDDSEFQ 240
 QY 241 LHQORLARERENPPQPKLGVAATPISARFQPKLTAVAESVLEGTDTTQSPKQSMKLG 300
 DB 241 LHQORLARERENPPQPKLGVAATPISARFQPKLTAVAESVLEGTDTTQSPKQSMKLG 300
 QY 301 GAGVTPPLAVTLDDKGLQAPDNPALNTLLKQTKDQTOHYLAHASSDGSQHLLDNKG 360
 DB 301 GAGVTPPLAVTLDDKGLQAPDNPALNTLLKQTKDQTOHYLAHASSDGSQHLLDNKG 360
 QY 361 HLPDIKSTATSYVLHNSHPGEIKGLAQAGTGSVSDGSGKISLGSQTQSHNKTMLSQ 420
 DB 361 HLPDIKSTATSYVLHNSHPGEIKGLAQAGTGSVSDGSGKISLGSQTQSHNKTMLSQ 420
 QY 421 PGEAHRSLLTG1WQHPAGARPOGESIRLHDDKTHILHPELGVWQSDKQTHSOLSRQAD 480
 DB 421 PGEAHRSLLTG1WQHPAGARPOGESIRLHDDKTHILHPELGVWQSDKQTHSOLSRQAD 480
 QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYVDVQGOVALTTPGRHKMSIMPSLDAS 540
 DB 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYVDVQGOVALTTPGRHKMSIMPSLDAS 540
 QY 541 PESHISLSLHFAHQAQGLLHGKSELEAQSVASHGRVLVADSEGLFSAAPKQGDGNE 600
 DB 541 PESHISLSLHFAHQAQGLLHGKSELEAQSVASHGRVLVADSEGLFSAAPKQGDGNE 600
 QY 601 KMKAMPQHALDEHFGHGHQISGFFHDDHQLNALVKNFRQOQACPLGNHOFHPGNL 660
 DB 601 KMKAMPQHALDEHFGHGHQISGFFHDDHQLNALVKNFRQOQACPLGNHOFHPGNL 660
 QY 661 DALVIDNOLGLHHTNPSPHETLDMGHLGSLAQKGLHYDQLTGKWTGAESDCKQLKKG 720
 DB 661 DALVIDNOLGLHHTNPSPHETLDMGHLGSLAQKGLHYDQLTGKWTGAESDCKQLKKG 720
 QY 721 LDGAAYLLKDGVEVKRLINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780
 DB 721 LDGAAYLLKDGVEVKRLINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780
 QY 781 AVIGVKNYALATEKGDIRSFOIKPGTOOLERPAOTLSREGISGELKOHVHDKHKNLYALT 840
 DB 781 AVIGVKNYALATEKGDIRSFOIKPGTOOLERPAOTLSREGISGELKOHVHDKHKNLYALT 840
 QY 841 HEGEVFHPRAWONGAESSWHKLALPOSESKLSLMSHEHKPIATFEDGSOHLKAG 900
 DB 841 HEGEVFHPRAWONGAESSWHKLALPOSESKLSLMSHEHKPIATFEDGSOHLKAG 900
 QY 901 GWHAYAAPERGLAVTSGSTQVFNRLMGVKGKVIPEGSLTVKLSAQTGMTGAERK 960
 DB 901 GWHAYAAPERGLAVTSGSTQVFNRLMGVKGKVIPEGSLTVKLSAQTGMTGAERK 960
 QY 961 SSKFSERTRAVAFNPTWSTPRPIKNAAYATOHGQWQREGKPLXYEMOQALIKOLDANVR 1020
 DB 961 SSKFSERTRAVAFNPTWSTPRPIKNAAYATOHGQWQREGKPLXYEMOQALIKOLDANVR 1020
 QY 1021 HNAPODQLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVKLSNGEIN 1080
 DB 1021 HNAPODQLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVKLSNGEIN 1080

QY 1081 SEPKSPFGKALVQSFNVRNSGQDLKSLQOAVHATPPSAESKLOSMLGHPVSAGVDMSHQ 1140
 DB 1081 SEPKSPFGKALVQSFNVRNSGQDLKSLQOAVHATPPSAESKLOSMLGHPVSAGVDMSHQ 1140
 QY 1141 KGEIPLGRORDPNDKTKALTKSLRLDVTIGELHELADRAKLYSDHKPDADQIKQLRQQF 1200
 DB 1141 KGEIPLGRORDPNDKTKALTKSLRLDVTIGELHELADRAKLYSDHKPDADQIKQLRQQF 1200
 QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260
 DB 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260
 QY 1261 SAEALAKKLTLLSLDSGESMSFSSRSYGGVSVFVPTLSKKYVPVPIPGAGITLDRAYN 1320
 DB 1261 SAEALAKKLTLLSLDSGESMSFSSRSYGGVSVFVPTLSKKYVPVPIPGAGITLDRAYN 1320
 QY 1321 LPSFRTSGGLNVSGFGRDGGVGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380
 DB 1321 LPSFRTSGGLNVSGFGRDGGVGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380
 QY 1381 RIGAAVSGTLQGTQLNSLKFPLTEDELPFIGHLTHTLTPAEILLQKGIHOMKQGSKLT 1440
 DB 1381 RIGAAVSGTLQGTQLNSLKFPLTEDELPFIGHLTHTLTPAEILLQKGIHOMKQGSKLT 1440
 QY 1441 FSVDTSANLDRAGINLINEDGSKPENGVTARVSAGLSASANLAAGSRERSTTSGOGFTTS 1500
 DB 1441 FSVDTSANLDRAGINLINEDGSKPENGVTARVSAGLSASANLAAGSRERSTTSGOGFTTS 1500
 QY 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
 DB 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
 QY 1561 SLELKRAEPTVSTNDISELTSTLGHKFKOSATTKMLAALKELDDAKPAEQHLHQHFSK 1620
 DB 1561 SLELKRAEPTVSTNDISELTSTLGHKFKOSATTKMLAALKELDDAKPAEQHLHQHFSK 1620
 QY 1621 DVVGDERYEAVALKLVIRQOQADSHMELGASHSTTYNNLSRINNDGIVELLKHFD 1680
 DB 1621 DVVGDERYEAVALKLVIRQOQADSHMELGASHSTTYNNLSRINNDGIVELLKHFD 1680
 QY 1681 AALPASSAKELGEMMNDPALKDIIKQLOSTPSSASVSMELKDGLREOTEKAILDGKVG 1740
 DB 1681 AALPASSAKELGEMMNDPALKDIIKQLOSTPSSASVSMELKDGLREOTEKAILDGKVG 1740
 QY 1741 REEVGVLFQDRNNLRVKSVSQSVSGSEGFNTPALLLGTSNSAAMSMERNIGTINPKY 1800
 DB 1741 REEVGVLFQDRNNLRVKSVSQSVSGSEGFNTPALLLGTSNSAAMSMERNIGTINPKY 1800
 QY 1801 QDQNTPRFRFTLEGGIAQANPQASALTDLKKEGLEMS 1838
 DB 1801 QDQNTPRFRFTLEGGIAQANPQASALTDLKKEGLEMS 1838

RESULT 2

US-09-431-614-8
 ; Sequence 8, Application US/09431614
 ; Patent No. 6624139
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Schading, Richard L.
 ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
 ; TITLE OF INVENTION: RESISTANCE
 ; FILE REFERENCE: 21829/41 (EBC-003)
 ; CURRENT APPLICATION NUMBER: US/09/431,614
 ; CURRENT FILING DATE: 1999-11-02
 ; EARLIER APPLICATION NUMBER: 60/107,243
 ; EARLIER FILING DATE: 1998-11-05
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: Patent in ver. 2.0
 ; SEQ ID NO 8
 ; LENGTH: 1838
 ; TYPE: PRT
 ; ORGANISM: Erwinia amylovora

US-09-431-614-8

Query Match 100.0%; Score 9448; DB 4; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MELKSLGTEHKAHVTAHNPVGHVALQOGSSSSPQNAASLASAEGKQKMPRIHQ	60
Db	1	MELKSLGTEHKAHVTAHNPVGHVALQOGSSSSPQNAASLASAEGKQKMPRIHQ	60
Qy	61	STAADGISAHQKQKSFSLGCLGTFKFSRPAQOGPTTHSGKATLRLDARDGETQH	120
Db	61	STAADGISAHQKQKSFSLGCLGTFKFSRPAQOGPTTHSGKATLRLDARDGETQH	120
Qy	121	EAAAPDAARLRSRGVGRNMDMAGRPMVKGSGEDKVPTQQRHQLNNFQMFQMTLS	180
Db	121	EAAAPDAARLRSRGVGRNMDMAGRPMVKGSGEDKVPTQQRHQLNNFQMFQMTLS	180
Qy	181	KWAHPASAGDRLOHSPPHIPGSHHEIKKEPVGTSKATTAHADRVHIAEDDDSEFQQ	240
Db	181	KWAHPASAGDRLOHSPPHIPGSHHEIKKEPVGTSKATTAHADRVHIAEDDDSEFQQ	240
Qy	241	LHQRLARERENPPQPKLGAVATPISARFPQKLTAVAESVLEGTDTTOSPLKPSQMLKGS	300
Db	241	LHQRLARERENPPQPKLGAVATPISARFPQKLTAVAESVLEGTDTTOSPLKPSQMLKGS	300
Qy	301	GAGVTPLAVTLDKGLQALAPNPPALNTLLKQTLGKDTQHYLAHHAASDGSQHLLDNKG	360
Db	301	GAGVTPLAVTLDKGLQALAPNPPALNTLLKQTLGKDTQHYLAHHAASDGSQHLLDNKG	360
Qy	361	HLFDIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDGSKGISLGSQTOSHNTMLSQ	420
Db	361	HLFDIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDGSKGISLGSQTOSHNTMLSQ	420
Qy	421	PGEAHRSLITGIWOHPAGAARPOGESIRLHDDKTHILHPELGVMSADKDTHSQLSROAD	480
Db	421	PGEAHRSLITGIWOHPAGAARPOGESIRLHDDKTHILHPELGVMSADKDTHSQLSROAD	480
Qy	481	GKLYALKDNRTLQNLSDNKSSEKLVDKISYSDVQGGVAILTDTGPRHKSIMPSLDAS	540
Db	481	GKLYALKDNRTLQNLSDNKSSEKLVDKISYSDVQGGVAILTDTGPRHKSIMPSLDAS	540
Qy	541	PESHISILHPADAHQGLHKGSELEAOSVAISHGRLVVADSEKGLFSAALPKQGDGNE	600
Db	541	PESHISILHPADAHQGLHKGSELEAOSVAISHGRLVVADSEKGLFSAALPKQGDGNE	600
Qy	601	KWKAMPOHALDEHFGHDHQSISGFPHDDHQLNALVKNNFRQOHCAPLGNDDHOFHPCWNL	660
Db	601	KWKAMPOHALDEHFGHDHQSISGFPHDDHQLNALVKNNFRQOHCAPLGNDDHOFHPCWNL	660
Qy	661	DALVIDNQLGLHHTNPEPHEILDHGLSLALQEGKLYHFDQLTKGWTGAESDCKQLKKG	720
Db	661	DALVIDNQLGLHHTNPEPHEILDHGLSLALQEGKLYHFDQLTKGWTGAESDCKQLKKG	720
Qy	721	LDGAAYLKQGEVRLINQSTSSIKGTENVFSLPHVRNKPBPQDALQGLNKDDKAQAM	780
Db	721	LDGAAYLKQGEVRLINQSTSSIKGTENVFSLPHVRNKPBPQDALQGLNKDDKAQAM	780
Qy	781	AVIGWKNYALATEKGDIRSFOIKPQTOLERPACTLSREGISGELKDIHVHDKONLYALT	840
Db	781	AVIGWKNYALATEKGDIRSFOIKPQTOLERPACTLSREGISGELKDIHVHDKONLYALT	840
Qy	841	HEGEVFFHPQREAWNGAESSSWHKLAIPQSESKLSLSDMSHEHKPIATFEDGSOHLKAG	900
Db	841	HEGEVFFHPQREAWNGAESSSWHKLAIPQSESKLSLSDMSHEHKPIATFEDGSOHLKAG	900
Qy	901	GWHAYAPERGLAVTSGSTVFENRLMQGVKQVIPSGLTVKLSAQTGMTGAERKVR	960
Db	901	GWHAYAPERGLAVTSGSTVFENRLMQGVKQVIPSGLTVKLSAQTGMTGAERKVR	960
Qy	961	SSKFSERIRAVAFNPTMSTRPIKNAAYATQHGQREGLKPLVEMQGALIKQIDAHVNR	1020
Db	961	SSKFSERIRAVAFNPTMSTRPIKNAAYATQHGQREGLKPLVEMQGALIKQIDAHVNR	1020

Qy	1021	HNAPODLQSKLETLLDGEHGAELNDMKRFRDELEQSATRSVTVLGQHGVKNSGEIN	1080
Db	1021	HNAPODLQSKLETLLDGEHGAELNDMKRFRDELEQSATRSVTVLGQHGVKNSGEIN	1080
Qy	1081	SEFKPSPKALVQSFNNVNRSGDLSKSLQQAQVHATPPSAESKLQSMGLHFVSAGVDMSHQ	1140
Db	1081	SEFKPSPKALVQSFNNVNRSGDLSKSLQQAQVHATPPSAESKLQSMGLHFVSAGVDMSHQ	1140
Qy	1141	KGEIPLGRQRPNDKALTAKSLRLILDVTITIGELHELADKAKLVSDHKPDADQIKQLRQOF	1200
Db	1141	KGEIPLGRQRPNDKALTAKSLRLILDVTITIGELHELADKAKLVSDHKPDADQIKQLRQOF	1200
Qy	1201	DTLREKRYESNPVHYTDMGFTTHNKALEANTDAVKAFINAFKKEHGVNLTTRTIVLESOG	1260
Db	1201	DTLREKRYESNPVHYTDMGFTTHNKALEANTDAVKAFINAFKKEHGVNLTTRTIVLESOG	1260
Qy	1261	SAELAKKLNTLLSLDGSMSFSRSGYGGVSTVPVPTLSKKVPVPVPGAGITLLDRAYN	1320
Db	1261	SAELAKKLNTLLSLDGSMSFSRSGYGGVSTVPVPTLSKKVPVPVPGAGITLLDRAYN	1320
Qy	1321	LSFRTSGGLNVSPGRDGVSGNTMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD	1380
Db	1321	LSFRTSGGLNVSPGRDGVSGNTMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD	1380
Qy	1381	RIGAAVSGTLQSTLQNSLKFKLTEDELPGFIHGLTHGLTTPAELLQKGIHQMKOGSKLT	1440
Db	1381	RIGAAVSGTLQSTLQNSLKFKLTEDELPGFIHGLTHGLTTPAELLQKGIHQMKOGSKLT	1440
Qy	1441	FSVDTSANLDIRAGINLNEGSKPNGVTARVSAGLSASANIAGSRERSTTSQFGSTTS	1500
Db	1441	FSVDTSANLDIRAGINLNEGSKPNGVTARVSAGLSASANIAGSRERSTTSQFGSTTS	1500
Qy	1501	ASNNRPFLNCGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALDNRTSOSI	1560
Db	1501	ASNNRPFLNCGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALDNRTSOSI	1560
Qy	1561	SLELKRAEPTVTSNDISELTSTLGHKFKDSATTMMALAKELDDAKPAQLHLOOHFSAK	1620
Db	1561	SLELKRAEPTVTSNDISELTSTLGHKFKDSATTMMALAKELDDAKPAQLHLOOHFSAK	1620
Qy	1621	DVGDERYEAVRNKLKLVIRQQAADSHSMELGSASHSTYNNLSRINNDGIVELLHKHFD	1680
Db	1621	DVGDERYEAVRNKLKLVIRQQAADSHSMELGSASHSTYNNLSRINNDGIVELLHKHFD	1680
Qy	1681	AALPASAKKLGEWMNDPALKDIIKQLOSTPPSSASVSMELKGLRQTEKAILDGKVG	1740
Db	1681	AALPASAKKLGEWMNDPALKDIIKQLOSTPPSSASVSMELKGLRQTEKAILDGKVG	1740
Qy	1741	REEVGVLFDORNRLRVKSVSVQSKSEGFNTALLGTSNSAAMSERNIGTINFKYG	1800
Db	1741	REEVGVLFDORNRLRVKSVSVQSKSEGFNTALLGTSNSAAMSERNIGTINFKYG	1800
Qy	1801	QDQNTPRRFTLEGGIAQNPQASALTDLKKEGLEMSK 1838	
Db	1801	QDQNTPRRFTLEGGIAQNPQASALTDLKKEGLEMSK 1838	

RESULT 3

US-09-350-852A-1
; Sequence 1, Application US/09350852A
; Patent No. 6596509
; GENERAL INFORMATION:
; APPLICANT: Bauer, David W.
; APPLICANT: Beer, Steven V.
; APPLICANT: Bogdanove, Adam J.
; APPLICANT: Collmer, Alan
; APPLICANT: Ham, Jong Hyun
; TITLE OF INVENTION: RECOMBINANT CONSTRUCTS AND SYSTEMS FOR SECRETION OF
; FILE REFERENCE: 19603/2061
; CURRENT APPLICATION NUMBER: US/09/350, 852A
; CURRENT FILING DATE: 1999-07-09

```
; PRIOR APPLICATION NUMBER: 60/092,357
; PRIOR FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-09-350-852A-1

Query Match      11.0%; Score 1041; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 7.2e-66;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSPQNAASIAAEGKRGKMPRIHQ 60
Db 1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSPQNAASIAAEGKRGKMPRIHQ 60

QY 61 STAADGISAHQOKKFSLRGCLGTTKFSRSPAQGPGTTHSKGATLRDLLARDGGETQH 120
Db 61 STAADGISAHQOKKFSLRGCLGTTKFSRSPAQGPGTTHSKGATLRDLLARDGGETQH 120

QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVYKGGSGEDKVPQOKRHQLNPFQGMQRTMLS 180
Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVYKGGSGEDKVPQOKRHQLNPFQGMQRTMLS 180

QY 181 KMAHPASANAGDRLOHSPPHI 201
Db 181 KMAHPASANAGDRLOHSPPHI 201

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match      2.7%; Score 257.5; DB 4; Length 10182;
Best Local Similarity 17.7%; Pred. No. 1.1e-07;
Matches 396; Conservative 319; Mismatches 805; Indels 723; Gaps 90;

QY 33 SSSSPQNAASL-AEGRKRGKMPRIHQPTAADGISAHQOKKFSLRGCLGTTKFSR 91
Db 6262 ATQSVQNAEQALHGAELNQDK---QTSSTELDGLTDLTDAORE-KLREQINTNSRDD 6316

QY 92 APQ--GPGTTHSKGATLRDLLARDG---ETOHEAAPDAAR----- 129
Db 6317 IKQIEQKALNDAMKLEQVAKQDGVHANSDYTNESQAQDAVNNALKAQEDINN 6376

QY 130 -----LTRSGGVKRRNMDMAGRPVYKGGSGEDKVPQOKRHQLNPFQGMQRTMLS 181
Db 6377 NPENLAQDITNALNNIKQAQNLHGAQKLIQ---QDKNTTNQAIGNLNLNPFQOKDALIQ 6432

QY 182 MAHPASA-----NAGDRLOHSPPHIPGSHHEIK----- 209
Db 6433 AINGATSRDQVAEKLEAEALDEAMKQLEDQVQDDQISNSPFINESDKQTYNDKIQ 6492
```

```
QY 210 --EPPVGSTSKAT-----TAHADRVELAQEDDDSEFQOLHQOBLARE 249
Db 6493 AAKEIINQTSNPTDKQKIADTLQNIKDAVNNLHGQD-KLAQSKQDANNQNLHDLDTTEE 6551

QY 250 RENPPQPKLGVAATPISARFOPKLTAVAESVLEGTDTTOSPLKPSQMLKSGAGVTPPLAV 309
Db 6552 QKN-----HFKP-----LINNADTRDEVNK----- 6571

QY 310 TLDKGKQLAPDNPPALNTLLKOTLGKDTQHYLAHASSDGSQHLLLDNKGHLFDIKSTA 369
Db 6572 -----OLEIAKOLNGDMSTLHKVINDKDOIQLHLSYNINADNDKKQYDN-----AIK 6618

QY 370 TSYSVLHNSHPGHIKLAQAGTGSVSDGSKSISLGSTQSHNKTMLSQPGEAHRSL 429
Db 6619 EAEEDLIHN-HPDTLDHKAQ-----DLNKKIDQAHNEL- 6650

QY 430 TGIWQHAPGAAR-----PQGESIR--LHDDKIHLHPGLGWQS 466
Db 6651 -----NGESRFKQALDNALNDIDSLSLNVFORQTVKDNINHVTTLESQAQEL----- 6698

QY 467 ADKDTSHSQRQADGKLYALKD-----NRTLQNLSDNKSEKLVDKI----- 508
Db 6699 -----QKAKELNDAMKARDSIMNQEQIRKNSNYTNEDLAQONAYNHAVDKINHILGE 6751

QY 509 -----KSYSDVQRCQVALTDTTPGRHKMSIMPSLD 538
Db 6752 DNATMDPQIIKQATODINTAINGLINGDQKIQDAKTDAKQOITNFTGLTEPQKQAL----- 6806

QY 539 ASPESHISLSLHFADAHQGLHGK---SELEASQVAISHGRLVVAD-----SEGLKF 587
Db 6807 ---ENIINQOTSANVAKQLSHAKFLNGKMEELKVAVAKASLVQNSNYINEDVSEKEAY 6863

QY 588 SAAIPKQGDGNE-LKMKAMPQHALDEHFGHDHOISGFFHDDHGOALNALVKNNFQQAACP 646
Db 6864 EQAIAX---GQEIINSENPTISSTDIRTIQBIQNDAAEQNLHGQ-----NKLRAQAEI- 6913

QY 647 LGNDHGFHPGWNUTDALIVNQILGLHHTPEPEHILDMGHLGSLALQEGKLYHFDQLTGK 706
Db 6914 AKNEIQNLGSLNSAQITKLIQDITRTTKPAVTQKLEAKAIQNAQMOQLKQSIADK----- 6969

QY 707 WTGAESDCKQLKGLDGAAYLLKDGVEVKRLNIN-----QSTSSIKHGTENVFSLPHVN 760
Db 6970 -----DATLSSNVYNEDSEKKLAYDNVAVSOAQLINQNLNPTWIDISNIQAITQ 7018

QY 761 KP-EPGDALQGLNKDDKAQAMAVIGNKYLALTEKGD-----IRSFQIKPGTQOLBERPA 813
Db 7019 KVIQAKDSLHGANKLAQNAQDSNLIINQSTNLNDKQKQALNDLINHAQTKQQAIEIIQA 7078

QY 814 QTLREGISGELKDIDVD---HKQNLIALTHEGEVHQPRAWQNGAESSSW----- 862
Db 7079 NKLNNE--MGTLLTVVEEQSNVHQOSKY-INEDPQVQNIYNDISIQKREILNGTDDVLN 7135

QY 863 -HKALPQSESKLSLDMSHKPIATFEDGSQH-----QLKAGWGHAAYAAPERG 911
Db 7136 NKKIADAIQNLHTKNDLHGDKQKLAQQAQDATNELNVLTNLNSQROSEHDEINSAFRT 7195

QY 912 PLAVGTSGSTQVFNRLMGVKGKVIKPGSLGTVKLSAQGTGGTGAEGRKVSKSFERRIAY 971
Db 7196 EVSNDLNHAKAL-NEAMRQLENEVALENSVK-KLSDFIN-----EDEAAQNEVSNALQKA 7248

QY 972 --AENPTWST---PRPKNAAYATOHQWQREGIKPIYEMQGALIKOLD---AHN----- 1018
Db 7249 KDIINGVPSSTLDKATTIEDALLEQNARESLHGQKLEQAKNAQVAIEDNLQALNPGQVL 7308

QY 1019 -----VRHNAPOPDLQSKLETLDLGEHGAELINDMKFRDELEQASATRSVTVLGQHGV 1073
Db 7309 AEKTLVNAQSTKPEQVEAL-----QKAKELNEAMKALKTEINK-----EQIK 7351

QY 1074 KSNGEINSEFKPSGPKALVOSFN--VNRSGQ-----DLKSLQQAQVHATPPSAESKL 1123
Db 7352 ADSRYVNAD-----SGLOANYNSALNYGSIITATTPPELNKQVINRATQTIKTAENNL 7405

QY 1124 --QSMLGHPVSAG-VDSHSHQGEIPLGFRQDPNDKALTLSRLILDTVTIGELHELADKA 1180
```


QY 797 IRSFQIKPGTQOLRPAQTLSREGISGELKDIHVDHKQNYALYTHGEVHFQPREAWQNG 856
Db 1393 ITKEGINAGSKQITNVADGINAKDAVNSQDLNLAQKN---ATDDAAVKYD-----1441
QY 857 ABSSSHKIALPOSE-SKLKSLDMSHEHPKPIATFEDGSO-HQLKAG-----GHWAAPE 909
Db 1442 -DAKTKDKVTLKGGDKGTVDNVKAGHSSTSEAVNGSQIHKISNKSIGGNTVVPND 1500
QY 910 RGPIA---VGTSGSQTIVNELMQ-----GVKQKVIPIGSLTVKLSAQTGMTGABGRKV 960
Db 1501 -GSLTNNIGGTGKNINDAISEVKNATKAKTTVTGDMNIVVKETVNDKGSNTYE---V 1556
QY 961 SSFESERIRAYAFNPTM--STPRPIKAAVATOHGMOGREGKLPYEMQALIKQLDAHN 1018
Db 1557 ATKKDLTNSVTTGDTVNNGLTIKDGPSITKDGINA-SGKQITNVADG--INAKDAVN 1613
QY 1019 VRH-----NAPQPDLOSKELET--DLGHEGABE--INDMKR-----1050
Db 1614 VDQITVKENLNGRITDTNNQLNDAKKDLGNQIADTNKNLNDAKKDLGNQITDTNTKINT 1673
QY 1051 PRDEQSATRSVTVLQHGVLKSGEINSEKPKSPGKALVQSFN--VNESGODLSKSLQ 1109
Db 1674 TKDQITQIINDTKELNNTGNTKT--ELMSKIDSTTELENKLNAGNSGNDVHRKLG 1731
QY 1110 QAVH-----ATPPSABSKLQSMGLHFVAGVDMSHQK---GEIPLGRQDPNDKALT 1159
Db 1732 EKNIIGGAARASTPAKTSGENVITRTTQDGIQIELLKDSKFDSTVTG-----NTLN 1784
QY 1160 KSRIL---DITVTIGELHELADKAKLVSDHKPDADQIKQIRQPDFTTIREKYESNPVKHY 1216
Db 1785 TNGLTKEGSPVTKGINAGKKTINVDAGVNAKDAVN--KSQDLNLAQKONATDDAAVK 1842
QY 1217 TDMGFTINKALEANYDAVKAFAKKEHGVNLTTRTVLESQSGAEAKKLK-----NT 1271
Db 1843 YDDAKTKDKVTLKKGOG--TWLDNVKAGH--ISSTSEAVNGSQIHNSIKNSIGGNT 1898
QY 1272 LLSLDSGESMFSRSGVSTV-----FVPTLSKVPVPVPIGAGITLDRAYNLSPST 1326
Db 1899 VVNPFG--SLTNNIGGTGKNINDAISEVKNATKAKTTTTEGDNIVVKETVN-----1950
QY 1327 SGGLNVSPRGDGVSGNIMVATGHDVM--PWTGKKSAGN-----ASDWLSAKH 1374
Db 1951 -----KDG--STNVEVSTKDLTNSVTGDTVLNNGLTIKDGPSITKDGVNAGG 1999
QY 1375 KISPLRIGA-----AVSGTLQTLQNSLKFKLITELDEPGFIHGLTHGLTTPAELLQ 1427
Db 2000 KKITDVANGVTAQNSKDAVGAQVHHISNIK-----NSIGGNT--VVNP-----2042
QY 1428 GIEHQMGQSKLITFSDVTSANLDRAGINLNEGSKPVGTVARVSAGLSASANLAAGSRE 1487
Db 2043 -----DGLSTNNIGGTGKNINDAIK-SVDEKVTGVDNLTQKGLNFGAN-----D 2088
QY 1488 RSTTSG-----QFGSTTSASNNRPTFLNGVAGANL-----TAALGVASHSTHEKFPVG 1536
Db 2089 QKTTQKGAVHRKLGDT-----INIVG-GADAKTAEDKTSGENIITRTTDEGVKIE 2137
QY 1537 IFP--APTSTNVSAALALDNRTSQSISLELKAEPVTGNDIS-----ELTSTL-GHKFKDS 1589
Db 2138 MLKDVKFSVNVG-----GHLVNOGLIIGKGPSITVNGINAGKQITNVADGINAKDA 2191
QY 1590 ATTKMLAALKELDAPQBLHLOQHFSAKDVVDGERYEAVRNKKLVIRQQAADSHSM 1649
Db 2192 VNKQQLD--KQINEVK--DQI-----GKD-IGKLSHAQVQYDKD---KNGVDKNSV 2235
QY 1650 ELGSASHSTYNNLSRINDGIVELLHGFDAALPASSAKRLGEMNNDPALKDIKQLQ 1709
Db 2236 TLGGGEGKPNLKNVA---DG-----KVAEGSKDAVNGGQLWNQONQVKNNDIKNIQ 2285
QY 1710 ST-----PFSSASVSMELKDKGLRETEKQKAILDKGVGREV 1744
Db 2286 NNIDNINSKAGLVQOQKPNGETVTKDTGGTINSMASKEG--DRVVGQVKDGEI-----2338
QY 1745 GVLFDQRNNLRVKS---VSVSQSVKSGEGNT---PALLIGTNSAAMSERN-----IGT 1794

Db 2339 -----XAGSNOAVNGGQIHKISSEKNSIGGNTTIDPKDSITNNIGGTGKNINDAIGT 2394
QY 1795 INFKYQDQNTPRRFTLEGGLAQANPQVASALTDL 1829
Db 2395 LN-----QSNQELGNKIINL 2409
RESULT 6
US-09-543-681A-4998
; Sequence 4998, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4998
; LENGTH: 1589
; TYPE: PRI
; ORGANISM: Proteus mirabilis
US-09-543-681A-4998
Query Match 2.5%; Score 238; DB 4; Length 1589;
Best Local Similarity 19.5%; Pred. No. 1,3e-07;
Matches 341; Conservative 238; Mismatches 603; Indels 570; Gaps 89;
QY 289 SPLKPSQMLKSGAGVTPPLAVTLDKGKLQAPDNPPLNTLLKQTL---GKTOHYL-AH 344
Db 172 NPLFENGQLKGYSTLNNTNLLSLGKGL-----NTGLLDLIAPRIDSRGKITAAEISAF 226
QY 345 HASDGSQHLHLLDNKGHLFDIKSTATSYSVLHNSHPGEIKG-----KLAQA 390
Db 227 TQNTFFSQH-----FDILSQKPVSAALDSVFFSGMQSGRIRIINTASGSGVKLAGK 277
QY 391 GTGSVSVVDGSKGKISLGSQTQ--SHNKTMLSQGEAHRSLLTGIWQHPAGARP-----442
Db 278 FTANDLSVRADNLTQDSQVRYDSYDK---DGSENYQNYRGGITVNNSSGSQLTKTEL 333
QY 443 QGESIRL---HD-----DKIHILHPELGV-----WQ--SA 467
Db 334 GKKNITLVASSHNQIKASDLMGDDITLQGADLTIDGKLOQKETDIDNRWFYSWKYDVTK 393
QY 468 DKDTHSQLSRQADGK---LYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQRGQVAIIT 523
Db 394 EKEQITQIGSQIDAKNNATLTATKGDVTLDAKINAGNNLAINANKDIHING-----LI 447
QY 524 DTPGRHKMSIMPSLDASPEHSIS--LSLHPADAHQGLHKGSELBAQSAVISHGLRVVAD 581
Db 448 EKESR-----SENGKRNHTSHLESGSWNSHQTTETLKASelta-----GKDLGLD 493
QY 582 SEGKLFSAAIKPGQDGNELKMKAMPQHALD-BHFCHDHQISGFPHDDH-----CQLNA 633
Db 494 AQGSITAGCAKLIHANENVL-VNAKDNINLNQKTNNDKTVT---DNHVMWGIGGGQN--547
QY 634 LVKNNFRQO---HACPLGNDHQFHPGWNLTDAVLDNOLGHLHTNPEPHEILDMHGLSL 690
Db 548 ---KNNNNQOQVSHATQLTADGQL-----LLAADNNVNI-----578
QY 691 ALQEGKLYFDQLTKEGWTCAESDCKQKKGDLGAAYLLKDGVEKVLN-INQSTSSIKGT 749
Db 579 -----TGSQ-----VKGNQGAFAVKTQGDVVDNAMSETISKIDERT 615
QY 750 ENVFSLPHVRNKPFG-DALQGLNKDDKAQAMAVIG--VNKYLALTEKGD-----IRSF 800
Db 616 GTAFNITKSSHKNETKQSTSELSISDAQLTVVSGNDVNNVLSLKSADKLGISLGD 675
QY 801 QIKPGTQ--QLERPAQTLSREGISGELKD-----IHVDHKQNYALYTHGEVHFQPREA 852

```

Db      676 NVKSAQVYKIDDEKTSLATGAKEVEDKQYAGFHITHTTKNTSTE-----724
QY      853 WONGAESMSHKLALPQSESKLSLMSHSHKPIATPEDGSOHQHAKAGWHAYAAPERP 912
Db      725 -----TEQANSTISGANVDLQANKNVIFA-GSDLKTTAGN-----758
QY      913 LAVGTSQSQTVE-----NRLMOGVKGVKIPGSGUTVVKLSAOTGMGTAEGKVSKEFSERI 968
Db      759 --ASITGDNVAFVSTENKKGQTDNTDTTISGG-----FSYTGQVD-----KVGSKAD---802
QY      969 RAYAFNPTMTSTPRIPKAAVATO-----HGWQREGLKELYEMOGA 1009
Db      803 --FOYDKQHTQTEVTKNRSQTEVAGDLTITANKDLLHEGASHVEGR-----YQESG-853
QY      1010 LIKQDLAHVNRHNPAPDQSKLETLDLG-EHGAELNDMKRPRDELEQSASTRSVTVLQ 1068
Db      854 -----ENIQHLAVNDSETSKTSLNVGIDGVNL-----D 883
QY      1069 HOGVLKSGEINSEFPKSPGKALVQSFNVRSGQ---DLKSGLQAAVHATPPSAESKLQSM 1126
Db      884 YSGVTK-----PVKKAIEDGVNTTKPGNNTDLTKV-----TARDAIANLANL 926
QY      1127 LG-HFVSAGVDMSHOKGEIPLGRQDPNDKTAITKSLIDVTIGELHELADKAKLVD 1185
Db      927 SNLETPNVGVEIGKGG---GSQKSTDSQAVSTS-----INAGKIN-----965
QY      1186 HKPDADQIKQLRQFPDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKXEH 1245
Db      966 -----IDSNKKLH--DQG-TH-----YQSTQ 983
QY      1246 HGVNLTTRTVLESQGSABLAKKLKNTLLSLDGSMSFSRSYGGVSTVFVPTLSKKVPV 1305
Db      984 EGISUTANT-----HTSEVAQDKHOT-----TFHETKGGG-QVGVSTKGTGSDIT 1026
QY      1306 PVIPIGAGITLDRAYNLSFRTSGGLNVSPGRDGVSGNIMVATGHDVMPYMTGKITSAGN 1365
Db      1027 VALKEGQTTNA--LMTKAKGSOFTS-----NGDISINVGEDA--HYEGAQFDAQ 1075
QY      1366 ASDMLSAKHKISPDRLIGAAVS--GTLQGTQLNSLKFKL-----TEDELPGFIHGLTH--1416
Db      1076 GKTVINA-----GGDLTLTAQATDTHESESQNVNGSANLKVITTPESKDYGGGNACTTHS 1131
QY      1417 -----GTLTPAELLQKGEIHKOMK-QGSKLTFVSDTSAN---LDLRAGINLNEDGSK 1463
Db      1132 KEQTTAKVCAITGSGOIELNAGHNLTLQGTLSSEQDIALNATNKVDLQSA--SSERTEK 1189
QY      1464 PNGVTARVSAGLS-----ASANLAAGSRERSTTSGQFSGTTSANNNPTFLN 1510
Db      1190 GNNLSGGVQAGFGKMTDDASSVNLGSAQFAIGKQDEKSVSREGGTTNNSGN-----1242
QY      1511 GVGAGANLTAALGVHASTH-EGKPVGIFPAFTSTNVSAALALDNR-TSQSISLELKRAE 1568
Db      1243 -----LTINGNSVHLOQAQVN-----SKUTOLTSQSGDIEITSQ 1277
QY      1569 PVTNDISELTST-LGHFKDASATTKMLAALKELDDAKAPQLHL-----QQHF 1617
Db      1278 ---STDYKNMGTDIGFNGKKTNTTP-----KEVTEKPATSIHNIIGKLLVNVEDQKT 1329
QY      1618 SAKDVWGDERYEARNLKLVII--RQQAADSHMSHLSGASHSTYNNLSRINNIGIVELL 1675
Db      1330 SHONATLETGTLTINSKDLTSLGANVTADSVTGNVGSLSNIAQSKESNRHVTYVGVN 1388
QY      1676 HKHFDALPASS-----AKRLGEMMNDPALKDI-----KQLQSTPPSSASVSMELK 1723
Db      1389 YNHTND--PKSSQVKNKATAKAGSL--EKTIKOTIDSGIKSTDAISDKYNSLSTIADK 1444
QY      1724 DGLBEQTEKALD--GRKV-----REEVGVLFDNRNLRKVSVSQSV 1765
Db      1445 TGISDET-KAKIDQFGKVGNGIKNIIVTGAEGHTANADIKVTHVD-NDAVTKTTLSTSN 1502
QY      1766 SKSEGFNTPALLTGTSNAAWSMERNTGTFNFKVQ---DONTPTRTLEGG---IAQANP 1820

```

```

Db      1503 DLSLVNNGSTKLTCAE-----IASKQGOVDLGGSSVKLENIGHHVEAGADL 1549
QY      1821 QVASALTDLKKKE 1832
Db      1550 DLKSSVVDLAKQ 1561

RESULT 7
US-09-841-786-1
; Sequence 1, Application US/09841786
; Patent No. 6669940
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.
; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3241
; TYPE: PR1
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-1

```

```

Query Match      2.5%; Score 236; DB 4; Length 3241;
Best Local Similarity 18.9%; Pred. No. 5.6e-07;
Matches 416; Conservative 339; Mismatches 776; Indels 668; Gaps 117;

QY      9 EHKAHVHTAHPVGHVVALQOQSSSSPQAAAAAAGKNGKMPRIHQFSTADGIS 68
Db      155 DFKKALEAKHKGKVGNGIIPVDG-KVKIPLNPGSITVEGK-----INAVEGIG 202
QY      69 --AAHQKKKFSL-----RGCLGTKKFSRS-----APOG 95
Db      203 LYAADIRLKDTAILKTGITDFKNLWISDRINSGLTGLKATKTSKGDILLSAHDSPK 262
QY      96 QPGTTHSGATLR-----DLLARDGETQHEAAPDAARLTTRSGVVRN--140
Db      263 AMGNSTVGGKRIEYVVKNTYANIESDAVLEADGNIKTSAKATNGRFRFKKEGKETYTP 322
QY      141 --MDDMAGRPMVKGS--GEDKVPTQQRHQLNPNFGQMTWLSKVAHPASANAGDLRQH 196
Db      323 LSLSDVEASVRVNGKVGKNDITAEAK---NF--YDALVTYKLA-----KH 365
QY      197 SPHPHPGSHHEI-----KEEVPVGTSTKATTAHA---DRVEIAQEDDD 235
Db      366 SFSFVTGSIPIINLNGFLGLTSSKVVVIGDKAKVEATEGKANLHYSYGVRAATGAAATSP 425
QY      236 SEFQOLHQRL-----ARENPPOPPLKGVATPIARFQPKLTAVAESVLE 282
Db      426 LKITNLYEKANGKLLSIGAGYISAKNSN-----VTIEGVKSKGRADITSSENTID 479
QY      283 ---GT---DITTSPLKPSQMLKSGAGVTPPLAVTLQK-KLQ LAPDN---PPALNTL-329
Db      480 ASVSVGTMRDSNKVAL---SVLVTGENKS--SVKIAKAGAKVESETDDVNVVRSEAINSR 534
QY      330 --LKQTLGK-----DTQHYL-----AHHASSD-----G 350
Db      535 AAVKGGGLDSCNGVVAANISYNNASSRIDVDGYLHAKKRLNVEAHNITKNSVLQTSGLG 594
QY      351 SQHLLDN---KGHLFDIKSTATSYSLVHNSHPEIKGKLAQAGTGSVVD-----398
Db      595 TSPFMDHNVYESGHLKSLD-AIKQRFPGSDSVNBEIKNKLTLNLFVSGVSATIANHNSAS 653
QY      399 ---GSKGKISLGGTQSHNKTMLSQPCAEHRSLLTGTIWHQHPAGAAARQGESIRLHDDKIH 455

```

654 VAIGESRLS--SGVEGSSNVRALNEAQNLRATTSSG-----SVAVRKEKK 697
456 ILHPELVGWSADKOT--HSQISROADKLYALKDNRTLQNLSDNKSSEKLVDKI 508
698 KLIGNAAVFGYKNNASVTIADHAEL--VSEKIDINSENK-IEYKNSPKMAKSVIDKL 754
509 KSVSVQDQGVAILTDTTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLHGSSEL--- 565
755 E-----LTKRAFGETYT--PEYDPKDIESIEKLL---NAFSEKLDCKPELLN 798
566 -EAQSVAISHGRVLVADSEGLFSAALPKQGDGNELMKAMPCHALDEHGFHDHQSGLFF 624
799 GERWTIILPDG---TSKGTGATEIANYVQGMKKLEEK-LPK-----GFK 839
625 HDHGOINLVKN--NFRQOHACPLGNDHGFHPCWNLITDALVIDNQLGLHHTPEPEIL 682
840 AFSEG-LSGLIKETLNF--TGVGNYANFH--TFSS-----GANGE-RDVS 879
683 DMGHGLSLALQEGKHLHFDOLTKGWTGA-----ESDCKOLKK-----GLDGAAYLLK 729
880 SVGAVSWVQEN---YSKVSUG-KGAKLAACKDLNKAINKAETVNLVGNIGLARST 934
730 DGEV--KRLNINQSTSS-----IKHGTENVS-----LPHVR---NKPEGDALQGL 771
935 SGSVAGKLNVRQSKNSAIVEAKEKAELSGENINADALNRLFHVAGSFGSGGNAINGM 994
772 NK-----DDKAQAMAVGVNKLALTEKGDIRSFOIKP-----GTQQLERPAQT 815
995 GSYSGGISKARVSDDEAYLKA-----NKKIALNSKNITSVWNAAGSAGIGYKN----- 1043
816 LSREGISGELKDIHVDHK-----ONLYALTHEGEVFPQPR-----EAWQNGAES-- 860
1044 -AAVGVAVANDYDINSKASTEDNDEGQSKDKKDDVTVTAESLEVDATKTGTINSIS 1102
861 ---SWHKLALPQSSSKLSLDMSEHKP-----IATPEDGSQHL----- 897
1103 VAGGINKVGSPESEKPKS-----EERPEGFGKIGNKVDVSKNKITSDMSLSLTKINY 1157
898 -----KAGWHAYAA--PERGP-LAVGTSQSQTVENRL-----MOGVKGKVPISGL 941
1158 ISEGVKKAGNIPSNVSHPTDKGPSFSLGASGSVS--FNNIKETSAVDGVKIN-LKGANK 1215
942 TVKLSAQTCGCTGABGRKVSXFSF-----RIRAYAFNPTMS--TPRPKNA--- 986
1216 KVEVTSSSTFVGAWGSSALQWNIHIGSGNSNISAGLAGAAVNNIQSKTSAVKNSDIR 1275
987 -----AYATOHGWQ--GREGLKPLYEMOGALIKOLDAHNVHNAPOPDLOSKLETIDL- 1037
1276 NANKPKVNALSGGTQVAAGAGLEAVKESGGQSKSYLLGTSASINILVWNEVSAKSENNTVA 1335
1038 GEHGAELLN-DMKRFRDE-----LEQSATRSVTVLQGHQGVKNGEINSEFKPSGKA 1090
1336 GESESQKMDVDVTAQADTQVTCALNLOAGKNGTVGATVTVAKLNKNVNASI--SGGRY 1393
1091 LVQSNVNRSGQDLKSLOQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQKGEIPLGRQR 1150
1394 T-----NVNREAD--AKALLATTQVTAAVT-----TGTTISSGAGLGNVQCAVSNNK-I 1438
1151 DPNDKTALPKSLIILDTVTIGELHELADKAKULVSHKPPADQIKOLRQOQFDL---REKR 1207
1439 DNDVEASVDKSSI-----EGANEINVIADKVGSSD-----LAKEYQALLNGKOKK 1484
1208 YESNPVKHYTDMGFTHNKALE-----ANYD-----AVKAFINAPK 1242
1485 YLEDRGINTNGYTYTKEQLEKAKKKEGAVIVNAALSVAGTDSAGGVALAVNTVKNKFK 1544
1243 KEHHGVNLTATRTVLESQGSAAELAK-----LKNLTLSDSGSMSPFSRYGGGVSTVFP 1297
1545 AELSGSN-----KEAGEDKIHAKHVNVVEAKSSTVVWNAASGLAISKDAFSGMGSGAWQ 1597
1298 TUSKVPVPVPGAGITLDRAYNLSFRSTSGGLNVSGFRGGVSGNIMVATGHDVMPYMT 1357
1598 DLSNDTIKVDKGR-ISAD-SLNVNANNILGVNVA-----GTIAGSLSTAVG----- 1643

1358 GKKTSGNADWLSAKHKISPDRLRIGAAVSGTLOQTLONSLKEFLTEDELPGFHGIHTG 1417
1644 -----AAFANNTLHNK-----TSALITGTKNPPFPFGKNTKVN-----VQALNDS 1682
1418 TLT-----PAELLOKGIH--OMKOGSKLT-----FSVDTSANLDIR-----A 1453
1683 HITNVSAGGAASIKQAGIGMVSVNRGSDETEALVSDSEFEGVSSFNVDKQDKTINTIA 1742
1454 GINLNEGSKPNGVTARVS--AGLSASANLAAGSRERSTTSG---QFGSTTSANNRPTFL 1509
1743 G---NANGGKACGVGATVAHTNIGKQSVIAIVKNSKITTTANDQDRKNINVTAKDYTTMT 1799
1510 NGVGAGANLTAALGVHSHSTHEKPGVGPAPFTSTNVSAALALD---NRTSQSISLEKR 1566
1800 IAVGVGAKGASVQGSASTTLNKTVS--SHVDQTDIDDKLEBENNGNKEKANVNVLAEN 1857
1567 AEPVTSN-----DISELTSTLGKHFKDSATTMKLAALKELDDAKPAEQ 1609
1858 TSQVNTNATVLSGASQAAGVAVVANKITQNTSAHIKNSTQVNRNALVKS----- 1908
1610 LHILQOHFSADV--VGDERYEAVERNKLKLVIRQQAADSHSMELGSGASHSTYN-NLSRIN 1667
1909 ---KSHSSIKTIGIG-----AGVGAGAGVTVGSVAVNKIV 1940
1668 NDIIVELLHKKHFDALPASPASAKRLGEMMNDPALKDIIKOLQSTPFSS--ASVSME--- 1721
1941 NNTIAELNHAKITA-----KGNVGVITESDAVIANVAGTVSGVARAAGASTSVNEITG 1994
1722 ---LXDG---LREQTEKAI-LDGKVGREBEVGLFQDRNNLRV-KSVSVSQSVSKSEGF 1771
1995 STKAIYKDSVTIAKEETDDYITTOGVQDK---VVDKVFKNLNLNEDLSQKRKISNKKGF 2050
1772 NTPALLIGT-----SNSAAMSMERNIGTINPK--YGQDQ 1803
2051 VTNSSATHLKSLLANAAGSQAGVAGTVNINKVYGETE 2089

RESULT 8
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ IDS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 2.5%; Score 233; DB 4; Length 3696;
Best Local Similarity 19.0%; Pred. No. 1,1e-06;
Matches 351; Conservative 277; Mismatches 717; Indels 506; Gaps 86;

QY 274 TAVASVLEGDTTQSPKLPQSKMLKSGAGVTPPLAVTLIDKGLQLAPDN---PPALNTLL 330
DB 53 TAQSDGGLNT--SQSNPISSEETNTLSGTQVPS---STENKQTVNPNHNAQPIAINT-- 106
QY 331 KQTLKQDTHYL---AHHASDGSQHLLLDNKGHLFDIKSTATSVLHNSHPGEIKGL 387
DB 107 EBAESAQTASYTNINENNDTSDGLHVNPQAKHIEAQSEDTVNTHTNHSNSISIPENK- 165
QY 388 AQAGTGSVSDGKSGKISL--GSGTQSHNKNTMLSQPGEAHRSLITIGIWHQFAGARPQ- 444

493 QNLSDNKSEKLVKIKSYSDVQGOVAILITDFGRHKMSIMPSLDASPEHSISLSLHFA 552
 187 KEIRSSKDNV---KYRRTTASGHASTAVGAMSYAQGHFSNAGTRATAKSAVSLAVGLA 243
 553 DAHQG---LHKGSELEAQSVAISHGRVLVADSEK--LPSAAIPKQGDGNEUKMAMPO 607
 244 ATAEGQSTIAIGSDATSSSLGALAGAGTRAOLQGSIALGQGSVVTQSDNNSPAYTPNT 303
 608 HALDEFGHDHQSIFGPHDDHCOLNALVKNPFQOQHACPLGNDHQHFGHWNLDALVI-- 665
 304 QALDPKFOATNNTKA-----GELS--IGSNSIKRKLIING-----AGVNTKIDAVNVAQ 349
 666 -----DNQGLGHTNPEPHEILDGMHGLSIALQEGKLHVPDQL 703
 350 LEAVVWAKERRITFOGDDNSTDVKIGLDT-----LTIKGA--ETNAL 392
 704 TKGWTA--BSDCKQLKGLDGAAYLLKQGEVVRKLNINQSTSSIKHG-----TENVPSLP 756
 393 TDNNIGVVKADNSGLKVKL--AKTLNLTENVNTTLN--ATTIVKVGSSSTTAELSDS 449
 757 HVNKPPEGD-----ALQGLNKDKAQAAMAVIGVKNYLALTEKDIRSFQIKPGCTQQ 808
 450 LTTQPNQTSQSTKVTYGVNGVFTNNAETTAIGTR----- 488
 809 LERPAQTLREGISGELKDIHVDHKQNYALYTHEGEVHFQPREAWQNGAESSMHKLALP 868
 489 -----ITRDKI--GFARDGDVDEKQPY-----LD 511
 869 QSESKLSLMSHBEKPIATPEDQSOHQKAGGHVAAAPERGLAVGTSGSQTVFRLM 928
 512 KKQLKVGSAV-----TIDNG--IDAGN-----KKISNLAKGSSANDAVTIEQL 553
 929 QGVKGVIPQSGTLV-----KLSAQTGMT-----GAEGRKVSSKFSERIRA 970
 554 KAAKPTLNAGAGISVTPTEISVDKSGNVTAFTNIGVKTIELNSDG--TSDKES--VKG 609
 971 YAFNPTMTSTPRPI-----KNAAYATQH 992
 610 SGTNNSLVTAHSLASYLNEVNRATDSALQSFVKEEDDDANAITVAKDTTKNAGAVSIL 669
 993 GWQREGKLPYEMOGALIKOL-----DAHVNRHNAQPDLOS---KLET 1034
 670 KLGKNGLTATKKGDTVTGFLSQDSGLTIGKSTLNNDGLTVKDTNEIQVGANGIKFTN 729
 1035 LDLGHEGAELNDMKRFEDELEQASATRSVTVLQHQGVKNGEINSFKEPSPGKALYQS 1094
 730 VNGNPGGIANTRAKITRDKLFGAGS-----DGAVDTNKPYLQDKLQGVNVTN 780
 1095 FVNRSGQDLSKSLQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQKGEIPLGRQRPND 1154
 781 TGINAGGKAIT-----GLSPTLPSI-----ADQSSR--NIBLGNTIQDKD 818
 1155 KTALTKSLILD-----VTIGELHELADK-AKLIVSD- 1185
 819 KSNAAASINDILNTGNLKNNNPIDVSTYDIVDPANGNATTATVTHDTANKTSKVVDV 878
 1186 -----HKPDADQIKOLRQOPDITLREKRYESNPVKHYTDMGFTHNKALE---ANYDAVK 1235
 879 NVDDTTIHLTGDDNKKLGKVTTKLNKTSANGN-----TAINFVNSDEADALYNAKDIA 933
 1236 AFINAFKEHHGWNLTTRTVLESQGSABELAKKLKNTLLSLDSEGSMSFSRSGGVSTVF 1295
 934 ENLNTLAKEIHTKTGADTALQT---FTVKVDENNADANAITVQCKNANNQVNTL- 988
 1296 VPTLSKKVPVPIPGAGITLDRAYNLSPS-RTSGGLNV--SPGRGGVVS-----GNIMVA 1347
 989 ---TLKGE-----NGLNITKDKNGVTFTGINTSGLKAGKSTLNDGGLSIKNPTGSEQIQ 1040
 1348 TGHDMVPMYTKTKSAGNASDWSLAKHKISPLDRIGAAVSGTFLQTLQNSLKFKLTEDBL 1407
 1041 VGADGVKFA-----KYNNNGVVGAGIDGIT-----RITRDEI 1072
 1408 PGFIHGLTHGTLTPAB--LLOKIEHQMKQSGSKLTFSDVTSANL----DLRAGINLNEDGS 1462

1073 -GFTG--TWGSLDKSPHLSKOGIN---AGGKIT-----NIQSGEQIANSHTAVTGG 1119
 1463 KPNQVTARYSAGLSASANLAAGSRERSTTSQGFSTTSASNNRPTF-----L 1509
 1120 KIYDLKTELENKISSAKTAQNSLHFSVADEQGNFTVSNPYSSYDTSKTSVDITPAGE 1179
 1510 NGVAGANITAAALGVAHSSTHEKPGVIGIPAFPTSTVWVAALALDNRTSQSISLELKRAP 1569
 1180 NGITTKVN---KGVVRVIGIDOTKGLTTPKLTIVGNNGKIVIDISQNGQ----- 1224
 1570 VTSNDISELTSTLGHFKQSA---TTKMLAALKELDDAKPAEOLHILQOHFS-----AK 1620
 1225 ---NTITGUSNTLIANTVNDKGSVRTTEQGNIIKDEKTRAASIVDVLSAGFNQNGEAV 1281
 1621 DVVGDERYEAVRNKLKLVIRQQAADSHMSLGASHSTTYNNLSR-----INNDGIVE 1673
 1282 DFV--STYDIV-----NFDAGN---ATTAKVTYDDTSKTSKVVDVNDTDTIE 1324
 1674 LLHKHF-----DAALPASS--AKELGEMMNDPALKDIIK 1706
 1325 VKDKLGVKTTTLTSTGTGANKFALSQAATGDALVKAASDIVAHSNLTLSDGIQTAKGASQ 1384
 1707 QLQSTPSSASVS-----MELKDGRLREOTEKAILDQKVGREEV--GVLFQDR 1751
 1385 ANNSAGYVDADGNKIVYDSTDNKYQAKNDGTVDKTEKAKDKLVAAQATPDGTGLAQ-- 1442
 1752 NNLRVKSVSVSQSV---SKSEGFNTF--ALLGTSNSAAMSERN-----IGTIN---F 1797
 1443 --MNVKSVINKEQVNDANKQGINEDNAFVKGLEKAAASDNKTNAAVTVGDLNAVAOTPL 1500
 1798 KYGQDQNTPRR-----FTLEGGIAQAQNPVASALTD 1828
 1501 TFGADTGTAKLGETLTIKGGQDTDN-----KLTD 1531

RESULT 10

US-09-489-039A-7973
 ; Sequence 7973, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 7973
 ; LENGTH: 2680
 ; TYPE: PRN
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-7973

Query Match 2.2%; Score 211.5; DB 4; Length 2680;
 Best Local Similarity 18.2%; Pred. No. 2.2e-05;
 Matches 365; Conservative 255; Mismatches 709; Indels 675; Gaps 81;
 QY 198 ANAGDRLOHSPHPHSGSHHEIKKEPVGSTSKATTAHADRE---IAQEDDDSEFPQLHQ 243
 DB 532 AKPGDRLOHLPFSGISQSRITQAVNRRQITVTTAYSTPRECVWAVESNDLFLO---Q 588
 QY 244 ORLARERPNPPPKLGVATPISARFQPKLTAFAESVLEGTDTTQSPLKQSMKLGSGAG 303
 DB 589 YRVTGVKEN-----GDATLITGVSHDPKFAIDTGAID 624
 QY 304 VTPPLAVTLDKGLQLADPNPP-----ALNTLLKOTLKGDTQHY-----LAHHASSDG 350
 DB 625 QRPVSV-----LPAGNQSPDDIVITRSVVNQGIVSETMOMNNSVSGALAYEAQ--- 675
 QY 351 SQHLHLLDNKGHLFDI-KSTATSYSVLHNSHPGKIGKLAQAQGTGVSVDG-KSGKISLGS 408

```

Db 676 -----WRNDGWINPRSSSTTGF-----EVSG--IYAGRYLVRVRAINAAETSSG- 719
QY 409 GTQSHNKTMLSPQE-----AHRSLTGI---WOHPAGAARPOGESIR----- 448
Db 720 WAYSEKTLTGKVGEPFLAPALATSLVHGQVSVNEFFTGS-----GDTLRTLOYSKNQD 775
QY 449 -----LHD-----DKIHLHPELG----- 462
Db 776 GSAPMPLSDVAVPGKSYOQGLSMGAEPFYARLVDRLGNESFWTCWQGMASDNFDDYY 835
QY 463 -----VW-----QSADKTHSOLRQA----- 479
Db 836 ENLTDIAIKDTAAWERTORTISTQGIINTQOELEQTAEALKEAEDQAKQVSQDIDASA 895
QY 480 -----DKYALKDNRTLQNLSDNKS-----SEKLVDKIKSYVDQRCQVAI----- 521
Db 896 KSITADVDBKISAVNKTITDEITSVNEALDSGLAQANKGVQEAASAVADANKQIATVNS 955
QY 522 LTDTPGRHKMSIMPSLDASPEHSISLHFADAHQGLHGKSELBAQSAVSHGLVAVD 581
Db 956 LTDSTIVORQSV---TDTAAEAENATIDLEIARVSKTLADGDAALNAQIKTAENGLKQSL 1012
QY 582 SEGKLFSAAIKQGDGNEKMKAMPQHALLDHFHGHQHSIGFFHDDHGLNALVKNFRQ 641
Db 1013 QVNTLTNAVQKQETADRIADVNAKASQAADLELLAATQGEIASIESLTQVMKTDENLARE 1072
QY 642 QHACPLGNDHQFHPGWNLTDAVIDNQLGLHHTNPEPHEILLDMGHLGLALQEGKLHYFD 701
Db 1073 MSSLAAGANICF-----DSQVI-----WHFNN 1094
QY 702 QLTGKWTGAESDCKQLKG-----LDGAAY-----LLK----- 729
Db 1095 QTTGKWTGAGVPGVSODGWLRLPADSATDPYTTPGGLAVDGAAYRFTMLRFRKTGKPV 1154
QY 730 DGEVKRLNINOSTSIIKHGTENVSFLPHVRNKPBGDQALQGLKDDKQAQAVIGVKYL 789
Db 1155 AGEIRWVSAGENFNNTKR-----YIVAEPEYADGVATLTVRDIPWTKNIDRIIR--L 1203
QY 790 ALTEKGDIRSQIKGTQQLERPAQTLSREGISGELKHVDHQNLYALTHEGEVEFHP 849
Db 1204 DLTNQODASNF-IEFDVIAGRPAGAS---TAALQDVRSTLSN---ALTAEA----- 1249
QY 850 REAWQNGAESSSWHKLALPOSEKLS-LMSHHEKPIATFEDQSOLKAGGHWAYAAP 908
Db 1250 -----QARSTLAAQMRGSDGSDLEKVTSGLLYQEKARTARVTAISAE 1290
QY 909 ERGPLAVGTSGSOTVFNRLMGVKGKVIKPGSLTVKLSAQT---CGMTGAEGRKVS----- 961
Db 1291 VK-----ARESLOTQFNDKAAVSGEL---SSLTQESQASRIGGLTSLGKKADAAL 1342
QY 962 SKFSERIRAYAFNPTM--STPRPIKNAAYATQHWQGR---EGLKPLYEMOGALIKQLD 1015
Db 1343 TSLTQKVEQOQATLTSQAALTSLTNRVQGTETGLAGTNEALSGLQSVTVQGDRI----- 1398
QY 1016 AHNVRNAPQDQLSKLETLDLGHGAILNDMKRFRDELEQSATRSVTVLGQHQGVLS 1075
Db 1399 -----TSQGSITKL-TSLLGTNAAL-----AKKAEAAVATLTQO---VEQ 1437
QY 1076 NGEINSEFKSPGKALVQSFNVRNSGQDLSKSLQAVHATPPSAESKLQMLGHFVSAGV 1135
Db 1438 NGR-----DIRSNTDSITSLNQLVNGOPNWRSLRYP-----V 1471
QY 1136 DMHQKGEIPLGRQRPNDKLTALTKSLRLDVTITIGELHELADKALVSKHKPAD----- 1191
Db 1472 QLAN-ACTVP-----SFSDVPAVPTV---VDEVADAAL--DFTSAGSVLIA 1513
QY 1192 -----QIKQLRQDFTUREKYESNPVKHYTDMG-----THNKALEANY 1231
Db 1514 LYSQCVKVAADTTITL-----APGARVFDGTGALFVNGQVAVGNASWNTVSPFLKAGW 1567
QY 1232 DAVKAFINAFKBBHH---GVNLTR-----TVLES 1258

```

```

RESULT 11
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

Db 1568 NTVEFLVNWQTCQAYINLGLKLDKVAEMYSGLVSAALAAAGVLSSNVSYQIGNEVVNS 1627
QY 1259 QGSAAELAKKLNKTLTSLDGSGESMSFSRSGYV-----STVEVPILSKKVPVFPV 1309
Db 1628 QBITQRLNALTOTDANVASKADQOTMNSLTGRVEKTESGLTAANANITSLKSAVRAGNAS 1687
QY 1310 GAGI-----TLDRAVN-LGFSRTS-----GGLNVSPGRDGGVSGNIMVATGHVD 1352
Db 1688 GGDLPNPTFPAYDQMGFSVVTTAEVPPSCPFGYAAARIASRDHHPFAFPATLNDV 1747
QY 1353 MPTWTKKTSAGNASDMLSAKHKISPDRLRIGAAV-----SGTQGTQNLNKLKFLTE 1404
Db 1748 IEISALVACGAGTANFNLYVGTAVRPDTSAGPLMAGGKSPSATWQRT---TWRFKVTQ 1804
QY 1405 DELP-GFI-----HGLTHGT-----LTPAELLQKGIHQMKQSKLTFSDVT 1445
Db 1805 AMVDGRTIRPFIQISQNSPYGTVFWFVTDWHRNVTAAQKQVDTADATAAADDLSLTTVTQ 1864
QY 1446 SANLDRAGINLINEDGSKPENGVTARVSAGLASANLAAGSRERSSTSGQFGSTTSASNNR 1505
Db 1865 QGNLLTSTG---NRITTOLENGU-ATTNAAVAKAD-ATAVQDLTNTVTQLGNDLTAANS 1919
QY 1506 PTFELNG-----VGAGANLTAALGVASHSTHECKPV----- 1535
Db 1920 ITKLTGNLANTDKALAKADATATLTDKVTQOGKTLESQNSLTLNLSLSQVAADD 1979
QY 1536 --GIFPAFTSTNVSAALADNRTSQSISLELKRAEPVTSNDISELTSTLGHKFKDSATTK 1593
Db 1980 ASGQIPGNLVNPSFERGLDGTGRS-----TATSVVEV-----SAPHS 2018
QY 1594 MLAALK-ELDDAKPABOLHILQOHFSAKDVVDERYEAVRNKLKLVIRQQAADSHSMELG 1652
Db 2019 GTRALKVDPCGSVSPQIIPVQ-----GRTELGWVWK-----EPG 2054
QY 1653 SASHTYNNLSRINNOGIVELHKKHFD-----AALPASSAKRL- 1691
Db 2055 ATTONGAGNNKLKRICNSAGQVFERPYNSGTVGNTWTLVSGRWKATETASLPVTLNLYI 2114
QY 1692 -GEMMNNDPALKKLIKQLOSTPFPSSASVSMELKDLREOTEKAILDKGVGREEVGVLPQD 1750
Db 2115 SGRYFDDFYVTDVTRVDIDATAGAVTG-----LTSRVSTAEGAITSQS 2159
QY 1751 -----RNNLRKVSVSQS-----VSKSEGTPTALLLGTSSAAMSMBERNI 1792
Db 2160 QQLTNLQNSLTNNSVSKADATALTSDVNRVTEASGKLTQSQQLTNLANVLTATRNA 2219
QY 1793 GTINFKYQDQONTFRFTLEGIA 1816
Db 2220 G-----DNLIPNDFLQGSTA 2235

```

APPLICATION NUMBER: US/08/409,995

FILING DATE: 24-MAR-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-61053/RTF

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1912 amino acids

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: unknown

US-08-409-995-4

Query Match 2.2% Score 209.5; DB 1; Length 1912;

Best Local Similarity 18.8%; Pred. No. 1.8e-05; Indels 615; Gaps 84; Matches 346; Conservative 221; Mismatches 655

QY 236 SEFQQLHQOARERENPPQPKLGVAIPISARFPQKLTAVASVLEGTDTTOSPLKQ 295
 DB 20 SELTRTHTKEL-RNRGDPV-----LATLLFATVQANATDEDEE-LDPVVRVAPVLSFHS 71
 QY 296 MLKSGS-----AGVPL-AVTLDKG---KLQAPD---NPPALNTLLKQ 332
 DB 72 DKEGTGEKTEVNSNMGYFDNKGVLKAGAITLKAGDNLKKQKXDTXTNASSFTYSLKK 131
 QY 333 TLGKDTCHYLAAHASSDGSQHLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGT 392
 DB 132 DLTDLT-----SVATEKLSFGANGDKVDITSDANGL-----KLAKTGN 169
 QY 393 GSVSVSDG-----KSGKISLGSQGTQSH-NKTMLSQPGEAHRSLLTGIWHPAGAAR 441
 DB 170 GNVHLNGLDSTLPDAVNTGVLSSSSFTPNDEVKTRAA---TVKDVNLNAGW-NIKAKT 224
 QY 442 PQG--ESIRLHD--DKHILHPELGVQWSADKDT-----HSQLSR 477
 DB 225 AGNVESVDLSAYANNVEFI-----TGDKNTLDVVLVTAKENKKTVEKFTPKTSVIK 276
 QY 478 QADGKLVALKDNRTLQNLSDNKSSEKLVDKISYVDQGVQVAILTDTGGRHMSIMPSL 537
 DB 277 EKGKLTGCKENNTNKVTSNATDN-----TD-EGNGLVTAKAVI 316
 QY 538 DASPESHISLHFADAHQGLHKGKSELEASVAISHGLRVVADSEGLFSAALPKQGDG 597
 DB 317 DAVNKAGWRVKTITANGQNG-----DFATVASGNTVTFES-GDGTASVTKOTNG 365
 QY 598 NELKMKAMPQHALDEHFHGDHQISQFFHDDHQLNALVKNFRQOHACPLGNDHGFHFG 657
 DB 366 NGITVK-----YDAKVGDLKDFD--- 383
 QY 658 NLTDALVIDNQLGHLHHTNPEPHEIL-----DMGHLGSLALQEGKLYFDQITKGWTG-A 710
 DB 384 ---SDKKIVADTALTITVGGKVAEIAKEDDKKLVNAGDLVTALGNL-----SWKAKA 433
 QY 711 ESDCKQLKGLDGAAYLLKQGEV-----KRLNINOS-----TSSIKHGTENVFSL 755
 DB 434 EADTDGALGIS-----KQGEVKAGETVTFKAGKNLVKQDQGANFTYSLQDALGLTGI 487
 QY 756 PHVRNKPEPGDALQINGKDDKAQAWA-----VIGVKNLYALTEKGDIRSFQIKPGTQQLER 811
 DB 488 TLGGTNGNDAKTIVNKDGLTITPAGNGTGTGNT-LSVYKDG-----IKANKAITN 540
 QY 812 PAQTL-----SREGISGELKDI--HVD-----HKQNLXALTHEGEVFFHQ 849
 DB 541 VASGLRAYDDANFDVNLNNSATDLNRHVEDAYKGLLENKANKQPLVTDSTAATVGDILR 600
 QY 850 REAW-----QNGAESSS-----WHKALPQSESLSKSLDM 879

DB 601 KLGVVVSTKNGTKBESNQVQADEVLFCTGAGATVTSKSENGKHTITVSAETK---ADC 657
 QY 880 SHEHKPIATFEDGSQHQKLAGWGHAYAAPERGLAVGTSGSTVFNRLMQGVKGKVI PGS 939
 DB 658 GLE-----KDGDTIKLV-----DNQNTDNLVTVGNNGTAVTKG 691
 QY 940 GL-TVKLSAQTGGMTCAGRGKVSSEKSEIRIRAYAFNPTMSTPRPIKNAVAATQHGQGRE 998
 DB 692 GPETVKT-----GATDADRGKV-----VKDATANDADKKVATVKDVATAINSA- 735
 QY 999 GLKPLYEMOGALIKOLDADNRHNAPOPLOS-----KLETLDLGEHGAELNDMKRF 1051
 DB 736 ---ATFVKTENLTSSIDENPTDNGKDDALKAGDTLTFKAGKNLVKVRGDKNITFDLAKN 792
 QY 1052 RDELEQSATRSVTVLGQHQGVLSKNGEINSEFKPSPGKALVQGFVNNRSQDL----- 1104
 DB 793 LEVKTAKVSDTLTIGN-----TPTGTTATPKVNTSTADGLNFAKETA 837
 QY 1105 ---SKSLQ---AVHATPPSABSKLOSMLGHFVSAGVDMSHQKGEIPL----- 1146
 DB 838 DASGSKNVYLKGIATTLTTFPSAGAK-----SSHVDLNVDAITKKSNAASIEDVLRAGWNIQ 892
 QY 1147 --GRQD-----PNDKTALTKSRLILDVTI-----GELHELADKAKLVSDHK 1187
 DB 893 GNGNVYVATYDTVNFDDSTGTT-----TVTVTKADGKGVKIGAKTSVVKDHN 945
 QY 1188 PDADQIKQLRQOQDFTLREKRYESNPVKHYTDMGFTHNKALEANVDAKAFINAFKE--- 1244
 DB 946 GKLTGKOLKD-----ANNGATVSEDDG---KDTGTGLVTAKTVIDAVNKSGWR 991
 QY 1245 --HHGNLTTRTVLESOGSAELAKUKNTLLSLDGSMSFSRSYGGVSVTFVPTLSKK 1302
 DB 992 VTGEGATAETGATAVNAGNAE-----TVTSGTSVNFK---NGNATVATV- 1032
 QY 1303 VPVPVPGAGITLDRAYNLSFRTSGLRVSGFDRGGVSGNIMVATGHDVMP-----YMT 1357
 DB 1033 -----SKDNGNINVKY--DVNVGDLKIGDKDKKI VADTTTLTVT 1069
 QY 1358 GKTSAGNASDWLSAKHKISPDLRIGAAVSGTGTQGLTQNSLKFKLTEDELPFGFHLTHG 1417
 DB 1070 GGVSVVPAGANSVNNKKL-----VNAEGLATALNLSWTAKADK---YADGESEB 1117
 QY 1418 TLTPAEILLQGIHQMGKSGKLTFSVDTSANLDRAGINLNEGSK-----PNGVTAR 1470
 DB 1118 E-----TDOEVKAGDKVTF-----KAGKNLVKQSEKDFYTSLODTLTGL 1157
 QY 1471 VSAGLSASANLAAGRSRSTTSQGFSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTH 1530
 DB 1158 TSITLGGTAN--GRNDTGTIVNKDGLTITLAN-----GAAAGTDSNGNTISVTK 1205
 QY 1531 EGPVGIIPAPTSTNVSAAL-----ALDNRTSQSISLELKRAEPV-----TSNDISE 1577
 DB 1206 DGISAG--NKEITNVKSALTKYKDTONTADETQDKEFHAAVKNANEVYFVGKNGATVSA 1262
 QY 1578 LTSILGKHFK--DSATTKMLAALKELDDAK-----PAEQHLILQQHFS 1618
 DB 1263 KTDNNGKHTVIDVAEKVGVGLEKDDTKLKLVDNTDGNLLTVDAKASVAKGEFN 1322
 QY 1619 AKDVVGVDERYEAVERNKL---KLVIHQQAADSHSMELSGASHSTYNNLSRINNDGIVELL 1675
 DB 1323 A--VTTDATTAGTINANERGVVK-----GNGATATETDKKVVATVG----- 1364
 QY 1676 HKHFDAAIPASAKRLGEMMNDPALKDIILQLOSTPFSSASVSMELKDGUREQTEKAIL 1735
 DB 1365 ---DVAKAIADAATFFVKVENDDSATID-----DSPTDDGANDALKAKDTL---TLKAK 1412
 QY 1736 DGKVBREVGVLQDNRNLRVKSVSQSKSEGFNTPALLLGTSNAAAMSERNIGTI 1795
 DB 1413 NLKVRGKNITFALANDLSVKASVSDKLS-----LGT-NGKNVITSDTKGL 1460
 QY 1796 NF-----KYGODONTPRFTLEGGTAAQNPQVASALTD 1828
 DB 1461 NPAKSKTGDDAN-----IHLNG-----IATSLTD 1485

1471 USAGLSANLAAGSRERSTTSGQFGSTTSANNRPTFLNGVAGAGANLTAALGVARSSTH 1530
 1158 TSITLGGTAN---GRNDTGVINKOGLTITLAN-----GAAAGTDSNGNTISVTIK 1205
 1531 EGKPVGIFPAFTSTNVAAL-----ALDNRSTQSISLELKRAPPV-----TSNDISE 1577
 1206 DGISAG---NKEITNVKSALKTYKDTQNTADETDQKEPFAAVKNAVEFVPGNGATVSA 1262
 1578 LSTLQKHKF--DSATTMMLAALKELDDAK-----PAQLHILQOHFS 1618
 1263 KTDNNGKHTVTDVABAKVGDGLEKDTGKIKLVNDTNGNLLTVDATKGASVAKGEEN 1322
 1619 AKDVVGDREYAVNLK---KLVRQQAADSHMSLGSASHSTTYNNLSRINNDGIVELL 1675
 1323 A--VTTDATTAQGTNANERKGVVK-----GSGNATATETDKKVAIVG----- 1364
 1676 HKHFDALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKGLRQTEKAIL 1735
 1365 ----DVAKINDAATFVKVENDDSATID-----DSPTDGDANDALKAXDTL--TLKAGK 1412
 1736 DGKVGREVGVLFDQRNLRVKSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTI 1795
 1413 NLKVRKDGKNTIFALANDLSVKSATVSDKLS-----LGT-NGNKVNITSDTKGL 1460
 1796 NF---KYGODQNTPRFTLEGGIAQANPOVASALTD 1828
 1461 NFAKDSKTGDAN-----IHLNG-----IASTLTD 1485

RESULT 13
 US-09-540-236-3459
 ; Sequence 3459, Application US/09540236
 ; Patent No. 6673910
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Berton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2005-001
 ; CURRENT APPLICATION NUMBER: US/09/540.236
 ; CURRENT FILING DATE: 2000-04-04
 ; NUMBER OF SEQ ID NOS: 3840
 ; SEQ ID NO 3459
 ; LENGTH: 2142
 ; TYPE: PRT
 ; ORGANISM: M.catarrhalis
 US-09-540-236-3459

Query Match 2.2%; Score 207.5; DB 4; Length 2142;
 Best Local Similarity 18.3%; Pred. No. 3e-05;
 Matches 365; Conservative 292; Mismatches 763; Indels 573; Gaps 93;
 8 TEHKAHVHTAHPNVGHGVALQGGSSSSPQAAASLAAEGKNGKMPRIHPQSTAAAGI 67
 530 TEVKSIAADAANGQVDQSFINLETANNKSXYAATI-----KDLH-----GL 572
 68 SAAHQOKSFSRLGCLGCTKTSRPAQCPGTHSK-----GATLRDILLARDGGET 118
 573 S---QVLTFA-----GDTGTAKKLGBILTIKGAADTRETLDKNGV 613
 119 QHEAAAPDAARLTRSGVKKRRNMDMAGRPVYKGGSGEDKVPQOKRHLNPNFGWRQM 178
 614 KDNQ-----NGGGLQVLAKNLSGLDAV-----NTKLTATGKIQVNNNGNTAQILL 659
 179 LSKMAHPASANGDRLQHSPPHPOSH-----HEIK-EBPVGSTSKATTAH--ADRV 227
 660 -----GGGLTFTQPNIGGQNSGKTYGVGDVKTDSNGNTPIAGTITYIYKEKI 707
 228 ETAAQEDD-----DSFQOLHOORLARENPPOPEKLGIVAT--PISARFPQKL-TA 275
 708 GFAQDQSGLDSEKPYLDKDLKVGKVIYTDGINAGGKAITGLPSTLNTTYTAPGVHTA 767
 276 VAESVLEGTDTTQSP-----LKPQSMKGGAGVTPLAVTLDKGKLQALPDNPALN--- 327

768 LHGSTLSNDKKTAAASIAADVLNAGFNLEGNAGVDFVSY-----DTVNFINGNA 817
 328 TLLKOTLQKXDTQHYLAHASSDSQ--HLLLDN--KHLFDIKSTATSYS-----VLHNS 378
 818 TTAQVYDNTKTSK-VAYDVNVDTTTHLTGTNGKQKQIGVKTTLTTTSAKGNTPINFV 876
 379 HPEIKKGKLAQAGTGSVDGSGKSLSGSGTQSHNKTMLS-----QPCGAHRSLL 429
 877 KPGD--DDALVKASDIDKDNLTNLAGEIRTAGKTASTALQTFSITDEHGNHTVGNPSSYD 935
 430 TGIWQHPAGAARQPGESIRLHDDKIHLPELQWQADKTHSOLSRQADGKLYALKDN 489
 936 TSKTFTVTVPAGENGITTKVKNKV-----KVGIDQTKGLTTPKLVGSSNGKGI-- 987
 490 RTLQNLNDKSSKSLVDKIKSYSDVORGQVAILTDTTPGRHKMSIMPSLDSPSHLSL 549
 988 -----DSKGGQNTITGLSN-----TLTDATN-----ATTGHVS-- 1015
 550 HPADAHQGLLHGKSELEAQS--AISHLRLVADSEKLPFAAIPKQ--DGNELKMKAM 605
 1016 -----EQGLAQAGANKTRAASIGDVLNAGFNLQNGEAKDFVSYDTVNFIDGNATTAKVT 1071
 606 PQHALDEHFCHDQISGFPHDDHQQNALVKNFRQOHACPLGNNDHOFHPGWNITDALVI 665
 1072 -----YDDTKQISTVTYD-----VNVV-----NKTLEVTG 1096
 666 DNQLGLHHTNPEPHEILDMGHLSLALQEGKLVFDOLTK-----GWTGAESDCKQL 717
 1097 DKXGVKTTT-----LTTSANGNATKFAADGDA--L 1127
 718 KGLDGAAYL-LKDGVEVKRL--NINOSTSSIKH-----GTENVFSLPHVRNKPPEPDALQGL 771
 1128 VKASDIATHLSTLAGDITQAKGASQASSASYVDADGNKVIY----- 1169
 772 NKDDKAQAMAVIGVKNYKALTEKGDIRSFOIKPGTQ-----QLERPQATLSREGISGLK 826
 1170 DSTDK-----KYQVNEKQVD--KITKEVTDKLVAQAQTPDGLTAAQMVKSVIN 1217
 827 DIHYDHKQNLVALTHEGEVHFQPREAMQNGAESWHKIALPOSESKLKSLDMSHEKPI 886
 1218 KEQVNDANKKQGINED-----NAFVKGLEKAA-----SDNKTNAAVT-----V 1256
 887 ATPEDGSOHLKAGWHAYAPERPGLAVGSGSTVFNRLMQGVKGVIPGSGLTVKLS 946
 1257 GDNLAVAQTPLTTFAGDTGTTAKLGE--TLTTKGQOTDNKLTNNIGVAGTGTGTVKLA 1315
 947 AQTGGM--TGAEGRKVSKFSERIRAVAFNPTMTPTREIKNAAYATQHGQREGKLPILY 1004
 1316 KDLTNLSVNAAGTTRIDEKGISFVDA--NGQAKANTPVL SANGLINMGG-----KRIS 1365
 1005 BMOGALIKOLDAHNVNRHNAPOFDLOSLETLDLGEHGAELNDMKRPFDELEQSATRSVT 1064
 1366 NI-GAAVDNDNAVNFQ-----FNEVAKTVNNLNNQNSGAS 1401
 1065 ---VLGQHOG--VLKSNGEINSEKPSPK--ALVQSFNVNRSGO---DLKSLSQAAVH 1113
 1402 LPFVVYTDANGRPINGTDGKPKAKIRGADGKYTHANGVFVDKGNPITDADKLANLAH 1461
 1114 ATPPSAESKLSMLGHFVSAGVDMSHQKEIPLGRQDRPNDKTALTATKRLILDTVTIGEL 1173
 1462 KXPLDAGQVVASLGG-NSDAITLTNIKSTLP--QIATPNTG-----NNTNAGOA 1508
 1174 HELADAKILVSDHKDPDQIKOLROQFDTLREKRVESNPVHYTDMGTGTHNKALANYDA 1233
 1509 QSL---PSLSAAQQAQNAASVKDVLNVGNFNQTHNHNQVDFVKAYDTVNFVNGTG--ADITS 1563
 1234 VKAFINAFKKEHHGNLTATRTVL---ESQGSALAKK-----LKNLTL---SLDSGEMS 1282
 1564 VRSADGTMS-----NITVNTALATDDGCVNLIRAKOGKFKYKADLMENGSLKAGKSAS 1617
 1283 FSRSS-----YGGGVSTVFVPTLSKKV-----PVPVTPGAGITLDRAYNLSF 1323

Db 1618 DAKTPTGLSIVNPADKSTGDAVALNLSKAVFKSKDGTITTTTSSDGSISQKDNSSI 1677
 QY 1324 SRTSGGLNVSFGRDGVSGNIMVATGHDVPMYMTGKTSAGNASDWLSAKHKISPOLRIG 1383
 Db 1678 TLSKDLGVG---GKVISNVGKT-----KDTDAAN-----VQQLNEVRNLLGLG 1719
 QY 1384 AAVSGTGLQTLNLSKFLKTEDELPGFIHGLT---HGTL-----TPABELLOKIEHOMK 1434
 Db 1720 NAGNODAGNQVNIADIK--KDPNSSSSNRITVIRKAGTVLGGKGNNDTEKLTATG----- 1771
 QY 1435 QGSKLTFSDVTSANLDEL-RAGINLNEDEGSKPVGTVARVSAG--LSASANLAAGSRERSTTS 1492
 Db 1772 -GVQGVGDKDNGANDLSNVWTKQDGSKKALLATYNAAGQTNVLTNNPASAIDRINQ 1830
 QY 1493 GQFGSTTSASNNRPTF--LNGVAGANLTPAALGVVAHSTHEGKPVGIFPAFTSTNVSAAL 1550
 Db 1831 GIRFFHVNDGNQEPVVGQRNGIDSSASGKHSVAIGFOAKADGE-----AAV 1876
 QY 1551 ALDNET--SOSISLELAKRAEPTVNDISELTSTL--GKH-----FKDSATTKMLAALKELD 1602
 Db 1877 ALGROTQAGNOSIATG-DNAQATGQCSIAIGTGNVVGKHSAGIADPSTVK----- 1926
 QY 1603 DAKPAEQLHILQOHFSAKDVVDGDERYEAVRNLKLVIRQQAADSHSMELGSASHSTTYNN 1662
 Db 1927 -----ADNSYSVGNV-----NQFTDATQTDVFGVG-----NN 1953
 QY 1663 LSRINDGIVELHKKHFDAAIPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMEL 1722
 Db 1954 ITVTESNSV-----ALGNSAISAGTHAGTQAKKSDGTAGTITTAGATGVK--- 2000
 QY 1723 KDLREOTEKALLDGKVGREEVGLVFDNRNLRVKSVSQ--SVSKSGGFNTPALLGTS 1781
 Db 2001 --GFAGQT-----AVGAVSVGASGABR---RIGNVAAGEVSATSTDAVNGSOLYKATQ 2048
 QY 1782 NSAAAMERNIGTINFKYQDQON-----TPRFTLEGGLAQANFQ-- 1821
 Db 2049 SIANTNE--LDHRHONENKANAGISSAMAMASMPQAVIPGRSMVTGGIATHNGQG 2103
 QY 1822 -VASALTDLKEG 1833
 Db 2104 AVAVGLSKLSDNG 2116

RESULT 14
 US-08-968-685A-10
 ; Sequence 10, Application Us/08968685A
 ; Patent No. 6214981
 ; GENERAL INFORMATION:
 ; APPLICANT: TUCKER, KENNETH
 ; APPLICANT: PLOSILA, LAURA
 ; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
 ; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS LLP
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10036-2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/968,685A
 ; FILING DATE: No. 6214981ember 12, 1997
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Baldwin, Geraldine F.
 ; REGISTRATION NUMBER: 31,232
 ; REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 790-9090
 ; TELEFAX: (212) 869-8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2123 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-968-685A-10
 Query Match 2.2%; Score 205.5; DB 3; Length 2123;
 Best Local Similarity 18.5%; Pred. No. 4.1e-05;
 Matches 332; Conservative 251; Mismatches 608; Indels 601; Gaps 94;
 QY 388 AOAGTGSVSVGKSG-----KISLGS-----GTQS-----HNKTMLSQPCGAHRSLLT 430
 Db 65 AVAGIGISEADGGKGANARGDKSIAIGDIAQALGSQSIAGDNKIVHNSNNAN----- 119
 QY 431 GIWQHPAGAAPOGSIIRLHDDKIHILPELGVWQS---ADKDTHSQLS-----ROA 479
 Db 120 -----IGAKASGNESIAIGSDVLASGHASIAIGSDDLYLKKTETVQOISELLPIIRGOKA 173
 QY 480 DGKLYALKDNRTLQNLSDNKSSEKLVKIKSVSDQGOVALLTDT-----PGRHKMSIMP 535
 Db 174 LNDIYQLADT-NLQYRTHAGGHASTAVGAWSY-AKGHFSNAGFTRATAGTYSVLAVGL 231
 QY 536 SLDASPESHISL-----SLHFADAHQGLLHGKSELEA--QSVAIHSGRLVWADSECKLPSA 589
 Db 232 TATAKAASSIAVGSNAQAIQFAATVAG---GSTQVNLNRGIALGFG-----SQ 276
 QY 590 AIPQGDGNELKMKAMPOHALDEHFGHDHQTSGFFDDHGLNALVKNFRQOQHACPLGN 649
 Db 277 VLQKNDNVNAANVRA-----YAPDDNQIDNRYKATFKN----- 310
 QY 650 DHQFHPGNWLTDLVIDNQLGLHHTNPEPHEILDGMHLGS-----LALQEGKLIHFD 701
 Db 311 -----GATDFVSGNSNG---NDSIRKLIINVG-AGSADTDAVNVAQKEAVRLANR 358
 QY 702 QLTGWTGAESDCKQLKKGL-----DGAAYLLKDGVEKVRININQSTSSI 745
 Db 359 QIT--FKGDDSN-NRVEKGLKTLITGAGTSALTQHNIGVQVQDGLKVQLAETLSL 415
 QY 746 KHGTENVFSLPHVRNKPEPDGALQGLKNDKQAQAMAVIGNVKNYLAALTEK---GDIRSFQ 801
 Db 416 KMT-----TENLTANEK-----VTVGKTRLTDDKIGFTNDMNGID 451
 QY 802 -----IKPGTOOLER-PAQTLRSREGIS-GELKDHVDHKKQNIYALTHEGEVPHQ 849
 Db 452 ESKPYLDKDTGIHAGGQKITKLTAGVVDVDDAATYQGLKKVNTQTAESALQTF----- 503
 QY 849 PREAWQNGAESSSWHKLALPQSSSKLSLDSMSHEHKPIAT-----FEDGSQHLKAG 900
 Db 504 VKKVDKNGDA-----NDSKI---ITVKNKNKPDGTQVNTLKLKGENGVDTVTETN 551
 QY 901 GWHAYAAPERGPLAVGTSG-----SQTVFNRLMQ-GVKGKVI-----PGSGLT-- 942
 Db 552 GTVTFLGNQNNGLTVGNSTLNDGLSVKNTNSNKQIQVGADGITFTDTSNKPAGIENT 611
 QY 943 -----VKLSAQTGGM-----TG--AEGRKVSKSESERIRAYAFNPT----- 976
 Db 612 TRITROGIGFANNVTGSLDANKPRLTPTGTINAGKELTNVQS-----AINPATNGGOLDF 665
 QY 977 ---MSTPRPIKNAAYATQHGQWQREGKLPYEMQ-----GALIK--- 1012
 Db 666 MNRLSANTKSGSAAT-----IKDLYNLSQVPLTFAGDTGNTVKLGEILKVKG 716
 QY 1013 -----QLDAHNVRRHNAPOPD--LQSKLE-----TLDLGEHGAEL 1044
 Db 717 GKTTADDLTKNNGVGVADSTDNSLTVKLAKTSLDLDVAVNTKLTASDKVTVDSGNNTAKL 776

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:20:22 ; Search time 88.3207 Seconds
(without alignments)
6477.980 Million cell updates/sec

Title: US-09-596-784-2
Perfect score: 9448
Sequence: 1 MELKSLGTEHKAHVTAHN.....NPQVASALTDLKKEGLEMK 1838

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	9448	100.0	1838	9	US-09-879-248-8
2	9448	100.0	1838	15	US-10-441-736-8
3	325	3.4	1957	12	US-10-365-742-120
4	275	2.9	2514	12	US-10-282-122A-66121
5	275	2.9	2514	15	US-10-320-800-40
6	273.5	2.9	6713	12	US-10-282-122A-43811
7	267.5	2.8	2659	14	US-10-311-879-28
8	262	2.8	6281	9	US-09-815-242-12996
9	261	2.8	2799	12	US-10-282-122A-65564
10	258.5	2.7	6641	12	US-10-282-122A-70580
11	258.5	2.7	10203	16	US-10-661-809-23
12	257	2.7	5795	9	US-09-815-242-12610
13	256.5	2.7	2398	12	US-10-282-122A-70176
14	253	2.7	2045	12	US-10-282-122A-74463
15	248.5	2.6	3158	9	US-09-815-242-12611

16	247.5	2.6	2344	9	US-09-815-242-12713	Sequence 12713, A
17	245	2.6	2703	12	US-10-282-122A-66108	Sequence 66108, A
18	243.5	2.6	3225	16	US-10-408-765A-254	Sequence 254, App
19	242	2.6	2434	9	US-09-815-242-5835	Sequence 5835, Ap
20	241	2.6	1577	12	US-10-282-122A-69069	Sequence 69069, A
21	240.5	2.5	2086	9	US-09-815-242-5639	Sequence 5639, Ap
22	240.5	2.5	5560	12	US-10-263-929-142	Sequence 142, App
23	239	2.5	2712	12	US-10-282-122A-67070	Sequence 67070, A
24	238	2.5	2283	14	US-10-172-502-4	Sequence 4, Appli
25	237.5	2.5	1905	15	US-10-259-194A-86	Sequence 86, Appl
26	236	2.5	2271	12	US-10-282-122A-43924	Sequence 43924, A
27	236	2.5	3241	9	US-09-841-786-1	Sequence 1, Appli
28	236	2.5	3241	12	US-10-647-057-1	Sequence 1, Appli
29	234.5	2.5	2437	9	US-09-815-242-5834	Sequence 5834, Ap
30	232	2.5	1981	9	US-09-928-457-38	Sequence 38, Appl
31	230.5	2.4	1491	12	US-10-282-122A-65570	Sequence 65570, A
32	229.5	2.4	2273	12	US-10-282-122A-66115	Sequence 66115, A
33	229.5	2.4	4322	16	US-10-437-963-104793	Sequence 104793, A
34	228	2.4	3533	12	US-10-282-122A-70177	Sequence 70177, A
35	228	2.4	3975	16	US-10-437-963-165014	Sequence 165014, A
36	227.5	2.4	2541	14	US-10-177-293-470	Sequence 470, App
37	227	2.4	3286	12	US-10-282-122A-49697	Sequence 49697, A
38	226.5	2.4	2025	9	US-09-815-242-5703	Sequence 5703, Ap
39	225	2.4	3692	12	US-10-282-122A-71235	Sequence 71235, A
40	223.5	2.4	1479	16	US-10-437-963-106122	Sequence 106122, A
41	223.5	2.4	2124	15	US-10-369-493-11841	Sequence 11841, A
42	222	2.3	4823	15	US-10-051-874-169	Sequence 169, App
43	221.5	2.3	5262	15	US-10-051-874-167	Sequence 167, App
44	221	2.3	2265	12	US-10-282-122A-45123	Sequence 45123, A
45	220.5	2.3	4952	15	US-10-051-874-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-879-248-8
; Sequence 8, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; PRIOR FILING DATE: 2001-06-12
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1838
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-09-879-248-8

Query Match	100.0%;	Score 9448;	DB 9;	Length 1838;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1838;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY	1	MELKSLGTEHKAHVTAHN	PVGHVALQOGSSSSPQNAASLAEGKRGKMPRIHQ	60
Db	1	MELKSLGTEHKAHVTAHN	PVGHVALQOGSSSSPQNAASLAEGKRGKMPRIHQ	60
QY	61	STAADGISAHQKQKSFSLRGCLGKTKFSRQAPQGGTTHSKGATLRDLLARDDETQH	120	
Db	61	STAADGISAHQKQKSFSLRGCLGKTKFSRQAPQGGTTHSKGATLRDLLARDDETQH	120	
QY	121	EAAAPDAARLTRSGGKVRKRNMDMAGRPVYKGGSGEDKVPYTOQKRHLNFFGQMRQTMLS	180	
Db	121	EAAAPDAARLTRSGGKVRKRNMDMAGRPVYKGGSGEDKVPYTOQKRHLNFFGQMRQTMLS	180	

Thu Jul 8 09:12:14 2004

181 KMAHPASANAGRLQHSPPHIIQGSHEIKEEPVGSTSKATTAAHADRVEIAQDDDDSEFOQ 240
181 KMAHPASANAGRLQHSPPHIIQGSHEIKEEPVGSTSKATTAAHADRVEIAQDDDDSEFOQ 240
241 LHQORLARENPPOPKLGVAIPISARPOPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
241 LHQORLARENPPOPKLGVAIPISARPOPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
301 GAGVTPPLAVTLKGLQLOLAPDNPALNTLLKOTLGKDTQHYLAHASSDGSQHLLLDNKG 360
301 GAGVTPPLAVTLKGLQLOLAPDNPALNTLLKOTLGKDTQHYLAHASSDGSQHLLLDNKG 360
361 HLFDIKSTATSYVLNHSHPGEIKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420
361 HLFDIKSTATSYVLNHSHPGEIKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420
421 PGSABRSLLTGIWOHPAGAAPPOGESIRLHDDKIHLHPELGVWQADKDTSHLSQROAD 480
421 PGSABRSLLTGIWOHPAGAAPPOGESIRLHDDKIHLHPELGVWQADKDTSHLSQROAD 480
481 GKLYALKDNRTQNLSDNKSSEKLVDKISYSDVQDQGOVAILTDTPCRHRKMSIMPSLDAS 540
481 GKLYALKDNRTQNLSDNKSSEKLVDKISYSDVQDQGOVAILTDTPCRHRKMSIMPSLDAS 540
541 PEGHISLSLHFAHAGHLLHGKSELEAQSVASISHERLVVADSEKLPFSAAIPOKQDGNEL 600
541 PEGHISLSLHFAHAGHLLHGKSELEAQSVASISHERLVVADSEKLPFSAAIPOKQDGNEL 600
601 KMKAMPOHALDEHFGHDHQTISGFFPHDHGQNLALVKNFRQOAHACPLGNDHQHPGWNLT 660
601 KMKAMPOHALDEHFGHDHQTISGFFPHDHGQNLALVKNFRQOAHACPLGNDHQHPGWNLT 660
661 DALVIDNOLGHHNTNPHEITLDMHGLGSLALQEGKLYHFDQLTGKWTGAESDCQKLLKG 720
661 DALVIDNOLGHHNTNPHEITLDMHGLGSLALQEGKLYHFDQLTGKWTGAESDCQKLLKG 720
721 LDGAAYLLKQGEVRLMINOSTSSI KHGTENVFLPHVRNKPPEGDALQGLNKDDKAQAM 780
721 LDGAAYLLKQGEVRLMINOSTSSI KHGTENVFLPHVRNKPPEGDALQGLNKDDKAQAM 780
781 AVIGVKNYALATEKGDTRSFQIKPGTOOLERPACTLSREGISGELKDIHVHDKONLYALT 840
781 AVIGVKNYALATEKGDTRSFQIKPGTOOLERPACTLSREGISGELKDIHVHDKONLYALT 840
841 HEGVEFHQPREAWONGAESSEWHKALPOSESKLSLDMSEHKPIATFEDGSOHLKAG 900
841 HEGVEFHQPREAWONGAESSEWHKALPOSESKLSLDMSEHKPIATFEDGSOHLKAG 900
901 GWHAYAAPERGLAVGTSGSTVFNRMLQGVKGKVIPGSGLTVKLSAQGTGWTGAERKV 960
901 GWHAYAAPERGLAVGTSGSTVFNRMLQGVKGKVIPGSGLTVKLSAQGTGWTGAERKV 960
961 SSKTSERIRAVAFNPTMTPRPKNAAYATOHGQREGKPLPYEMOGALLKOLDANVR 1020
961 SSKTSERIRAVAFNPTMTPRPKNAAYATOHGQREGKPLPYEMOGALLKOLDANVR 1020
1021 HNAPODLOQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGOHQGVLSKNGBIN 1080
1021 HNAPODLOQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGOHQGVLSKNGBIN 1080
1081 SEFKPSGKALVQSFNVNRSGDLSKSLQQAQVATPPSAESKQLQSMGLGHFVSAGVDMSHQ 1140
1081 SEFKPSGKALVQSFNVNRSGDLSKSLQQAQVATPPSAESKQLQSMGLGHFVSAGVDMSHQ 1140
1141 KGEIPLGRQDNDKTALTLSLLDVTITIGELHELADKALVSHKPDADQIKOLROOF 1200
1141 KGEIPLGRQDNDKTALTLSLLDVTITIGELHELADKALVSHKPDADQIKOLROOF 1200
1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGYNLTTRTVLESQG 1260
1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGYNLTTRTVLESQG 1260
1261 SAELAKKLKNTLLSLDSGSMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320

1261 SAELAKKLKNTLLSLDSGSMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
1321 LSFRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTKKTSGAGNASDMLSAKHKISPD 1380
1321 LSFRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTKKTSGAGNASDMLSAKHKISPD 1380
1381 RIGAAVSGTLOQTLONSLKFKLTDELPGFHGLTHGTLTPAELLQKGIHQMKQSKLT 1440
1381 RIGAAVSGTLOQTLONSLKFKLTDELPGFHGLTHGTLTPAELLQKGIHQMKQSKLT 1440
1441 FSVDTSANJDLRAGINLNEDEGSKPNVGTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500
1441 FSVDTSANJDLRAGINLNEDEGSKPNVGTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500
1501 ASNNRPTFLNGVGAGANLTAALGVVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
1501 ASNNRPTFLNGVGAGANLTAALGVVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
1561 SLELKAAEPVTSNDISLSTLTGLKHKFDSATTTKMLAALKELDDAKPAEQHLIIQOHPSAK 1620
1561 SLELKAAEPVTSNDISLSTLTGLKHKFDSATTTKMLAALKELDDAKPAEQHLIIQOHPSAK 1620
1621 DVVGDERYEAVRNLKLVIRQQAADSHMELGSAHSTTYNNLSRINNDGIVELLHKHFD 1680
1621 DVVGDERYEAVRNLKLVIRQQAADSHMELGSAHSTTYNNLSRINNDGIVELLHKHFD 1680
1681 AALPASSAKELGEMMNNDPALKDIIKQLOSTPFSSASVSMELKDGLREOTEKAILDKVG 1740
1681 AALPASSAKELGEMMNNDPALKDIIKQLOSTPFSSASVSMELKDGLREOTEKAILDKVG 1740
1741 REEVGULFQDRNLRVKSVSQSVSKSEGFTPALLLGTNSGAAMSMERNIGTINFKYG 1800
1741 REEVGULFQDRNLRVKSVSQSVSKSEGFTPALLLGTNSGAAMSMERNIGTINFKYG 1800
1801 QDQNTPRFTLEGGIAQANQVASALTDLKKEGLEMS 1838
1801 QDQNTPRFTLEGGIAQANQVASALTDLKKEGLEMS 1838

RESULT 2

US-10-441-736-8
; Sequence 8, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (EBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1838
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-10-441-736-8

Query Match 100.0%; Score 9448; DB 15; Length 1838;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELKSLGTEHKAHVTAHNPVGHVALQOQSSSSSPQNAASLAAGCKNRGKMPRIHOP 60
DB 1 MELKSLGTEHKAHVTAHNPVGHVALQOQSSSSSPQNAASLAAGCKNRGKMPRIHOP 60
QY 61 STAADGISAAHQKKSFSIRGLGCLGTKFKFSRGAPOQPGTTHSKGATLLDLARDDETQH 120

Db 61 STAAAGISAHQKKSFSURGLGKTKFSRSAPQPGTTHSKGATFLRDLLARDGGETQH 120
QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVPTQOKRHLNNFQOMRTMLS 180
Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVPTQOKRHLNNFQOMRTMLS 180
QY 181 KWAHPASANAGDRLOHSPPHIPGSHHEIKEEPVGSSTKATTAHADRVEIAQEDDDSEFQQ 240
Db 181 KWAHPASANAGDRLOHSPPHIPGSHHEIKEEPVGSSTKATTAHADRVEIAQEDDDSEFQQ 240
QY 241 LHQQRLARERNPPOPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPKPKQSMKLS 300
Db 241 LHQQRLARERNPPOPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPKPKQSMKLS 300
QY 301 GAGVTPFLAVTLDDKGLQAPNDPPALNTLLKQTKGDKTOHYLAHASSDGSQHLLLDNKG 360
Db 301 GAGVTPFLAVTLDDKGLQAPNDPPALNTLLKQTKGDKTOHYLAHASSDGSQHLLLDNKG 360
QY 361 HLFDTIKSTATSVLVNHPGIBKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420
Db 361 HLFDTIKSTATSVLVNHPGIBKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420
QY 421 PGEAHSLLTGIWHPAGAARPOGESIRLHDDKIHILHPELGVWQADKOTHSQLSROAD 480
Db 421 PGEAHSLLTGIWHPAGAARPOGESIRLHDDKIHILHPELGVWQADKOTHSQLSROAD 480
QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQDQGVAILTDTPGRHKMSIMPSLDAS 540
Db 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQDQGVAILTDTPGRHKMSIMPSLDAS 540
QY 541 PESHSISLHPADAHQGLLHOKSELEAQSVAISHGRLVVDSEGLFSAALPKQGDGNEL 600
Db 541 PESHSISLHPADAHQGLLHOKSELEAQSVAISHGRLVVDSEGLFSAALPKQGDGNEL 600
QY 601 KMKAMPQHALDEHGHCHDHOISGFHDDHGOALNALKVNNFRQHACPLGNDHFPGWNLT 660
Db 601 KMKAMPQHALDEHGHCHDHOISGFHDDHGOALNALKVNNFRQHACPLGNDHFPGWNLT 660
QY 661 DALVIDNQLGLHHTNPPEHIELDHGLGSLALQEGKLYFPQLTGKWTGAESDCKQLKKG 720
Db 661 DALVIDNQLGLHHTNPPEHIELDHGLGSLALQEGKLYFPQLTGKWTGAESDCKQLKKG 720
QY 721 LDGAAYLLKDGKVRKLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780
Db 721 LDGAAYLLKDGKVRKLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780
QY 781 AVIGVKNYALATEKGDIRSFOIKPGTOOLERPQOTLSREGISGELKDIHVHDKQNLALT 840
Db 781 AVIGVKNYALATEKGDIRSFOIKPGTOOLERPQOTLSREGISGELKDIHVHDKQNLALT 840
QY 841 HEGEVFHPQREAWNGAESSSHKLALPQSBKSLKSLDMSHEHKPIATFEDGSQHLKAG 900
Db 841 HEGEVFHPQREAWNGAESSSHKLALPQSBKSLKSLDMSHEHKPIATFEDGSQHLKAG 900
QY 901 GWHAYAAPERGLAVGTSGSTVFNRLMQGVKVPFSGGLTVKLSAQTCGTMGAERKV 960
Db 901 GWHAYAAPERGLAVGTSGSTVFNRLMQGVKVPFSGGLTVKLSAQTCGTMGAERKV 960
QY 961 SSKFSEIRIRAYAFNPTWSTPPIKNAAYATQHGQREGIKPLYEMQGALIKQDAHNVR 1020
Db 961 SSKFSEIRIRAYAFNPTWSTPPIKNAAYATQHGQREGIKPLYEMQGALIKQDAHNVR 1020
QY 1021 HNAFPQDLQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGHQGVLSKNGEIN 1080
Db 1021 HNAFPQDLQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGHQGVLSKNGEIN 1080
QY 1081 SEFKPSGKALVQSFNNVRSGQDLKSQQAQVHATPPSAESKLSMLGHFVSAGVDMSHQ 1140
Db 1081 SEFKPSGKALVQSFNNVRSGQDLKSQQAQVHATPPSAESKLSMLGHFVSAGVDMSHQ 1140
QY 1141 KGEIPLGRQDPNDKLTAKSRLLDVTITGELHELADKAKLVSDHKPDADQIKQLRQOF 1200

Db 1141 KGEIPLGRQDPNDKLTAKSRLLDVTITGELHELADKAKLVSDHKPDADQIKQLRQOF 1200
QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVAFINAFKKEHGVNLTTRTVLESQG 1260
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVAFINAFKKEHGVNLTTRTVLESQG 1260
QY 1261 SAELAKLKNLTLLSDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
Db 1261 SAELAKLKNLTLLSDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320
QY 1321 LSFSTSGGLNVSPORDGVGNGNIMVATCHDVMPYMTGKTSAGNASDWLSAKHKISPLD 1380
Db 1321 LSFSTSGGLNVSPORDGVGNGNIMVATCHDVMPYMTGKTSAGNASDWLSAKHKISPLD 1380
QY 1381 RIGAAVSCTLOGTLONSLKFKLITDELPGFTHGLTHGLTTPAELLQKGIHOMQOGSKUT 1440
Db 1381 RIGAAVSCTLOGTLONSLKFKLITDELPGFTHGLTHGLTTPAELLQKGIHOMQOGSKUT 1440
QY 1441 FSVDTSANLDRAGINLINEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
Db 1441 FSVDTSANLDRAGINLINEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
QY 1501 ASNNRPTFLNGVGAGANITAAALGVAHSTHGBKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
Db 1501 ASNNRPTFLNGVGAGANITAAALGVAHSTHGBKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
QY 1561 SLELKRAEPTVNSIDISELTSTLKGKFKDSATTKMLAALKEILDDAKPAQLHLOQHFSAK 1620
Db 1561 SLELKRAEPTVNSIDISELTSTLKGKFKDSATTKMLAALKEILDDAKPAQLHLOQHFSAK 1620
QY 1621 DVVGDERYEAVERNLIKLVIRQQAADSHSMELGSHASHSTTYNNLSRINNDGIVELLKHKFD 1680
Db 1621 DVVGDERYEAVERNLIKLVIRQQAADSHSMELGSHASHSTTYNNLSRINNDGIVELLKHKFD 1680
QY 1681 AALPASSAKRLGEMWNNNDPALKDIIKQLOSTPFSASVSMELKDGLEQTEKAILDGVKG 1740
Db 1681 AALPASSAKRLGEMWNNNDPALKDIIKQLOSTPFSASVSMELKDGLEQTEKAILDGVKG 1740
QY 1741 REEVGVLFQDRNNLRVKSUSQSVSKSEGTNTALLGTSNSAAMSMERNIGTINFKYG 1800
Db 1741 REEVGVLFQDRNNLRVKSUSQSVSKSEGTNTALLGTSNSAAMSMERNIGTINFKYG 1800
QY 1801 QDQNTPRRFTLEGGIAQANPOVASALTDLKKEGLEMSK 1838
Db 1801 QDQNTPRRFTLEGGIAQANPOVASALTDLKKEGLEMSK 1838

RESULT 3

US-10-365-742-120
; Sequence 120, Application US/10365742
; Publication No. US20030204868A1
; GENERAL INFORMATION:
; APPLICANT: Collier, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Cartinhour, Samuel W.
; APPLICANT: Schneider, David J.
; APPLICANT: Tang, Xiaoyan
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING
; TITLE OF INVENTION: NUCLEIC ACIDS, AND USE THEREOF
; FILE REFERENCE: 19603/4112
; CURRENT APPLICATION NUMBER: US/10/365,742
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/356,408
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/380,185
; PRIOR FILING DATE: 2002-05-10
; NUMBER OF SEQ ID NOS: 209
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 120
; LENGTH: 1957
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato DC3000
US-10-365-742-120

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 66121
LENGTH: 2514
TYPE: PRT
ORGANISM: Neisseria meningitidis
US-10-282-122A-66121

Query Match 2.9%; Score 275; DB 12; Length 2514;
Best Local Similarity 18.8%; Pred. No. 9.8e-10;
Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY 7 GTEH-KAAVHTAAHPV- - - - -GHCVALQQSSS- - - - -SSPQNAAS 43
Db 37 GSAHVKSVPFGTHAPVCRNIFSFLLGFSLCIAVGTANAFADGIIADKAAPKTOAT 96

QY 44 LAAGEKNRGMRIHQPTAAGDISAAHQOKKSFSLRGCLGTTKFRSAPQGP- - - - - 98
Db 97 ILQTGNG- - - - -IPQVNIQTPTSAGVSNQYQAFVGNRGAILNN- - - - -SRSTNTQTQLGGWIOG 151

QY 99 - - - - -TTTSHK- - - - -GATRLDOLLARDGETQHEAAAPDAABLTR 132
Db 152 NPWLARGEARVVVNQINSSHSOMNGYIEVGRRAEVVIANPAGIAVNGGFTNASRATL 211

QY 133 SGGVKRRNMDMAGRPMVK- - - - -GSGED- - - - -KVPTQQRKHQNLNFGQWRQTMLS 180
Db 212 TTQCPQVQAGLGSFKIRQGNVVIAGHLGDARDTDFTRILSYHSKIDAPVWGQDVRVAG 271

QY 181 KMAHPASANAGRLQHSPPHIPGSHHIEIKEEPVGVSTSKATTAAHADRVEIAQEDDDSEFQ 240
Db 272 QNDVVAATGNA- - - - -HSP- - - - -ILNNAANTSNNTANNTHIPLFAIDTG- - - - - 312

QY 241 LHOORLARENPPOPPLKLVGATPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
Db 313 - - - - -KLK- - - - -GMYANKITLISTA- - - - -EQAGIRNOGQLFAS 342

QY 301 GAGVTPPLAVTLDKGLQALADPNPALNTLLKTLGDKDQHYLAHAS- - - - -SDGSQHLLDN 358
Db 343 SGN- - - - -VAIDANGRLVNSGTMAAN- - - - -AKDTDNTABHKVNIRSQG- - - - -VEN 385

QY 359 KG- - - - -HLFDIKSTATSYS- - - - -VLHNSHPGEIKGKLAQAGTGSVSDVCSGKI 404
Db 386 SGTAVSQGTQIHQSQIQNTUTLLSSGEILHNS- - - - -GSLKNETSGTI- - - - -EAARL 435

QY 405 SLGSGTQSHNKTMLSQPGEAHRSLLITGIWQHPAGAARPGESIRLHDDKIHLHPGLVW 464
Db 436 AIDTDT-LNNQKLSQTG- - - - -SQKLHI- - - - - 458

QY 465 QSADKDTHSQLSRQADGKLYALKONRTLQNLSD- - - - -NKSSEKLVDKIKSYSDVQSGQ 518
Db 459 - - - - -DAQGKM- - - - -DNRGRMLQDPTAPTASNGSSNQTCN- - - - -SYNASPHSS 498

QY 519 VAILTDTGPRHKMSIMPSLDASPEHSISLSLHFDADAHQGLLHGKSELEAQSVASHGRLV 578
Db 499 TTTPTTATGTATVSNITAPT- - - - -FAD- - - - -GTIRTHGALDNSSSIANGOTD 547

QY 579 VADSEKGLFSAAIKQGGDNELKMKAMPQHALDEHFGHDHQSISGFFHDDHQLNALVKN- 637
Db 548 VSAQGG-LNNAG- - - - -QIDIHQLNAG- - - - -SAFDNHG- - - - -TIISDAVHIQAGSLNNQNGI 598

QY 638 NFRQCHACPLGNDHQFHPGNLTDALVID-NOLGLHHTNPEPHEILDMGHLSLALQSGK 696
Db 599 TTRQQ- - - - -LEITDQLDNAHGKLSABIAVLAVSSLLNNQNGE 638

QY 697 LHYPDQLTKGWTGAESDCKQLKGLDGAAYLLKDEVKRLNINOSTSIKHGTENVFSLP 756
Db 639 IATNQQL- - - - -IIHQGQOSTAVIDNNTGTIQSGROVAIOAK 675

QY 757 HVRNKPFGDALQGLNK- - - - -DDKACAMAVIGVKNYKALTEKGDIRSFOIKPCTOOLE 810
Db 676 SLSN- - - - -NGTLAADNKLDIALQDDFVVERNIVAGNE-LSLSTRGSLKNSHTLOAGKRIR 730

QY 811 RPAQTLSPRG- - - - -ISGELKDIHVDHKQNLVALTHEGEVHPQPREAMQNGAESSSWHKLA 866
Db 731 IKANNLDNAAQGNISQGGTTDGTQHN- - - - -LTNRGLIDGQ- - - - - 767

QY 867 LPOSESKLKSLSMDSH- - - - -EHPKIATFEDGSOHLKAGGWHAVAAEPGRPLAVGTS 918
Db 768 - - - - -QTKIQAGQWNNITGTGRIYGDNIATRLDNQDENGTA- - - - -ATAARENLMIG- - - - - 819

QY 919 GSOTVFNRLMQGVKVIPIGSLTVKLSAQTTGGTGAEGRKVKSKFSERIRAYAFNPMTS 978
Db 820 - - - - -QLNNRENSLIYSNDMAVGALDNGQATGKAQR- - - - - 853

QY 979 TPRIKNAAYATQGWQGRGLKPLYEMQGALLKQL- - - - -DAHVRNRNAP 1024
Db 854 - - - - -IHNAGATIEAAGKMRGLGVEKHLNTHLKTQLVETGREHIVDYEAFCRHELLREGT 909

QY 1025 QPDL- - - - -QSKLETLDLGEH- - - - -GAELLNDMK 1049
Db 910 QHELGSVYNDESDDLRTPDGAHWHYDYKVTQKTQVOTAPAKIISGNDLTIDGK 969

QY 1050 R-FRDELEQSATRSVTVLQHQH- - - - -VLKSNGEINSEPKPS-PGKALVQSFN 1096
Db 970 EVENTDSQIIAGNLIIVQTEKQGLHNEQTFEGKKVSENGKLSHYWEEKHGR- - - - -DSTG 1026

QY 1097 VNRSQDLSKSLQAVHATPPSAESKLQSMGLHFVSAVDMSHOKGPIPIGRORDPNDKT 1156
Db 1027 HSEQNTYLPBEEITRNISLGSFAYESHKALSHAPSQGTLPQSN- - - - -ISLPYTSNFT 1083

QY 1157 ALTKSRL-ILDTVTYICELHELADKAK- - - - -LVSDHKPDA- - - - -DQIKQLRQOQDTLREKR 1207
Db 1084 PLPSSSLIYINPVNKGVLVETDPRFANYRQWLGSQDYMVLSKLDLPNNLHKRGDGYEOR 1143

QY 1208 YESNPVKHTDMGFTHNKALEANYDAKAFINAFKKEHGVNLTTRTVLESQCSAEALAKK 1267
Db 1144 LINEQIAELT- - - - -GHRRLDGVQNDDEQFKALMDNGATAAASMLSVGIAL- - - - -SAEQVAQ 1197

QY 1268 LKNTLLSLSGSMSTSRSGGVGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSFRTS 1327
Db 1198 LTSDIWMLVQKE- - - - -VKLPDGGTQTVLVPQVYVRVKNGDIDGKAL- - - - -LSGSNTQ 1247

```

QY 1328 GGLVTSFGDGGVSGNIVATGHDVMPYMTGK---KTSA-----GNASDWLSAKHKIS 1377
D 1248 INVSGSLKNSGIIAGNALIINTDLDNIGRIHAQKSAVATQDINNIGMLSAQTL 1307
QY 1378 PDLRGAAY-SGTLOGLTQNSLKFTELDELPG-FHGLTHGTLT-----PAELLOKXIE 1430
D 1308 --LNAGNNINSOSTASSONTQSSSTYLDRMAGIYITGKEGVLAQAAGKDINIAGQIS 1365
QY 1431 HQMGKSKLTFSDVTSANLDL-----RAGINLNEDGSKFNGVTARYSAGLSASA----- 1479
D 1366 NOSEQG-QTRLOAGRINDLTVTSKHQATHFDADNHVIRGSTNEVGSSIQTKGDTVILS 1424
QY 1480 --NLAAGSRSTTSQGFSTTSASNRPFTFLNGVAGANLTAALGVVAHSSTHEGKPVGI 1537
D 1425 GNNLNKAAEVSSANGTL-AVSAKND---IN-ISAGINTT---HVVDAKSHTRSGG- 1473
QY 1538 PFAFTTNSAALALNDRTSQSISLELKRAEPVTSNDISELSTL-----CKHEK 1587
D 1474 ---GNKLVTITDKAOSHETHAQSSSTFEKGQVVLQAGNDANILGSNVISDNGTQOAGNHVR 1530
QY 1588 DSATTKMLAALKELDDAKPAEQHLILQOHPFSAKDVVVDERYEAVRNLKKLIVIRQQAADSH 1647
D 1531 IGT-----QTSQSETHTQKSGLSMAGIG-----FTI---GSKTN 1565
QY 1648 SMELGSASHSTTYNNLSRINNDGIVELLKHGFDALPASAKLGEHM-----NNDPALK 1702
D 1566 TQENQOSNEHTGSTVGS�KGDTTI-VAGKHYE-----QIGSTVSSPEGNNTIYAQ 1615
QY 1703 DIKQLOSTPFSSASVSMELKDLGTEKAILDKGVGREVG-----LFQDRNLIV 1756
D 1616 SIDIOAHNKLNSNTTQTYEQKGLTVAFSPVTD--LAQQAIAVAOSSKQVGOSKND-RV 1672
QY 1757 KSVSVQS--VSKSEGNTEPALLGTSNSAAMERNIGTFNKYQDQNTPRFT-LEG 1813
D 1673 NAMAANAVAGQAVTQKSAQNLANGTNAQVS-----ISITYGEQNN-RQTQVQA 1723
QY 1814 GIAQANPQVASALTDLKKEGLEMK 1838
D 1724 NQAQASQIQAGGKTTLIATGAEQS 1748

RESULT 5
US-10-320-800-40
; Sequence 40, Application US/10320800
; Publication No. US20030215469A1
; GENERAL INFORMATION:
; APPLICANT: ROBINSON, ANDREW
; APPLICANT: GORRINGE, ANDREW
; APPLICANT: HUDSON, MICHAEL
; APPLICANT: REDDIN, KAREN
; TITLE OF INVENTION: MULTICOMPONENT MENINGOCOCCAL VACCINE
; FILE REFERENCE: 1581.0790001
; CURRENT APPLICATION NUMBER: US/10/320.800
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: PCT/GB99/03626
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40
; LENGTH: 2514
; TYPE: PR
; ORGANISM: Neisseria meningitidis
US-10-320-800-40

Query Match 2.9%; Score 275; DB 15; Length 2514;
Best Local Similarity 18.8%; Pred. No. 9.8e-10;
Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY 7 GTEH-KAAVHTAAHNPV-----GHGVALQQGSSS-----SSPONAAS 43
D 37 GSAHVKSVPFGTTHAPVCRSNIFSFLLGSCLAVGANTAFADGIIADKAPKTOAT 96
QY 44 LAEGKRGKMPRIHQPSTAADGISAHQKKSFLRCCLGTGKFKESAPQQPG----- 98

```

```

D 97 ILQTGNG---IPQVNIQTPTSAGSVNQVQAQFDVGNRGAILNN--SRSTIQTQLGWIQ 151
QY 99 -----TTHSK-----GATLRLDLARDGETQHEAAAADAAALTR 132
D 152 NPWLARGEARVVVNQINSSSHSSQMGYIEVGGRRABVIANPAGIAVNGGGFINASRATL 211
QY 133 SGGVKRRNMDMAGRPWKV-----GSGED-----KVPTQOKRHQLNFFQGMQRTMLS 180
D 212 TTGQPOYQAGDLSGFKIRQGNVVIAGHGLDARDTDFRILSYHSKIDAPVWGQDVVAVG 271
QY 181 KMAHPASANAGDRLQHSPEHIFCSHHEIKEEPVGVSTSKATTAAHADRAVEIAQEDDDSEFOQ 240
D 272 QNDVWATGNA-----HSP-----ILNNAANTSNNTANNGTIPIPAIDTG----- 312
QY 241 LHQORLAREPREPPPKLGVATPIISARFPKLTAVAESVLEGTDTTQSPLEKQSMKGS 300
D 313 -----KLQ-----GMVANKITLISTA-----EQAGIRNQGLFAS 342
QY 301 GAGVTPPLAVTLKQKQLAPDNPPLAINTLLKQTLGKDTQHYLAHAS--SDGSQHLLLDN 358
D 343 SGN-----VAIDANGRLVNSGTMAAAN-----AKDONTAEHKVNIQSOG-----VEN 385
QY 359 KG-----HLFDIKSTATSYS-----VLHNSHPGEIKGKLAQAAGTGSVSDGKSGKI 404
D 386 SGTAVSQOQTQIHSQIQNTGTLSSGEILHNS-----GSLKNETSGTI-----EAARL 435
QY 405 SLGSGTQSHNKTMLSQFGEAHRSLLTGIWOHPAGAAPQGESIRLHDDKTHIILHPELVGW 464
D 436 AIDTDI-LNNQKLSQTG-----SOKLHI----- 458
QY 465 QSAKDQTHSQLSQADGKLYALDKNRTLQNLSD-----NKSSEKLVDKIKSYSDVQRG 518
D 459 -----DAQGKM-----DNRRMGLQDTAPTASNGSNQTCN-----SYNASPHSS 498
QY 519 VALTDPGRHKMSIMPSLDASPESHISLSLHFEADAHQGLLHGKSELEAQSVAISHGRV 578
D 499 TTTPTTATGTGTATVISNITAPT-----PAD-----GTIRTHGLDNGSSIANGQTD 547
QY 579 VADSEGLKFSAAIPKQDGNELKVKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN- 637
D 548 VSAQGG-LNAG--QIDIHQLNAG--SAFDNHNG--TIISDAVHIQAGSLNNQNGNI 598
QY 638 NFRQQAACPLGNDHQHPHGWNLTDALVID-NQGLHHTNPEPHEILDMHGLSLAQEGK 696
D 599 TTRQ-----LEIETDQLDNLNAHGKLLSAEIALAVSGSLNNQNGE 638
QY 697 LHYFDQLTKGWTGAESDCKQKGLDGAAYLLKDGGEVKRLNINOSTSSIKHGTENVFSLP 756
D 639 IATNQJ-----IHDGQOSTAVIDNTNGTIOGRDVAIOAK 675
QY 757 HVNRKPPEPGDALQGLNK-----DDKAQAMAVIGVKNKYALTAKGDIRSQIKPQTOOLE 810
D 676 SLNS-----NCTLAADNKLIDALQDDFYVERNI VAGNE-LSLSTGSLKNSHTLQAGKRIR 730
QY 811 RPAQTLREG-----ISGELKDIIHVDHKQNLIALTHEGEVPHOPREAWQNGAESSWHKLIA 866
D 731 IKANNLDNAAQGNIQSGGTTDGTQHN-----LTNRGLIDGQ----- 767
QY 867 LPQSEKLSLDMSH-----EHKPIATFEDGSQHLKAGGWHYAAAPERGPLAYGTS 918
D 768 -----QTKIQAGQMNIGTGRIGYDNIATRLDNQDENGTA--ATAARENLNIG-- 819
QY 919 GSQTVFNRLMQGVKQVPGSGLTVKLSAQTGMTGAEGRKVSFKSERIRAYAFNPTMS 978
D 820 -----QLNRENLSIYSGNDMAVGALDINGOATGKAQR----- 853
QY 979 TPRDIKAAATQHQWQGRGLKPLXEMOGALIKOL-----DAHNVRNAP 1024
D 854 -----IHNAGATIEAGKQRLGVEKLIHNTNEHLKTLQVETGREHIVDYAFGEHLLRGT 909
QY 1025 QPDI-----QSKLETDLIGEH-----GAELNDMK 1049

```

Db 910 QHELCHSVYNDSDHLRTPDGAHAENWHKVDYKVTQKTQVQTAPAKIISGNDLITDGG 969
 QY 1050 R-FRDELESATKSVTVLQHQH-----VLKSNGEINSEFKPS-PGKALVQSFN 1096
 Db 970 EVNTDSQIIAGNLIQVTEKDLHNEQTGFKVFSSEKLSHYWREKHGR---DSTG 1026
 QY 1097 VNRSQDLSKSLQAAVHATPPSAESKLSMLGHFVSAGVDMHSHQGEIPLGRQDPNDKT 1156
 Db 1027 HSEQNTLPEETRNISLSGFAYESHKALSHAPSQGTLPQSG---ISLPTNSFT 1083
 QY 1157 ALTKSRL-ILDTVTIGELHELADKAK-----LVSDHKPDA---DQIKQLRQFDTLREKR 1207
 Db 1084 PLPSSSLXIINPNKYLVEVTPRFANYRQWLGSDYMLDSLKLDPNLHKRLDGGYYEOR 1143
 QY 1208 YESNPVHYTDMGFTHINKALEANYDAVAFINAFKEHGHVNLTRTVLESQGSABELAKK 1267
 Db 1144 LINEQIAELT--CHRRLDGYQNDDEQKALMDNGATAARSMNLVSGIAL----SAEQVAQ 1197
 QY 1268 LKNVTLISLDSGSMFSRSYGGVSTVFPVTLSSKKVPVPIGAGITLDRVNLFSRSTS 1327
 Db 1198 LTSDIVMLVQKE-----VKLPDGGTQVLPVYVRYKNGDIDGKAL-----LGSNTQ 1247
 QY 1328 GGLNVSFGRDGVSGNIMVATGHDVYPMYTGK---KTS-----GNASDMLSAKHKL 1377
 Db 1248 INVSGSLKNSGTIAGNALINTDILNIGGRIHAQKSAVATQDINNIGGMLSAEQTL 1307
 QY 1378 PDLRTGAIV-SGTLOGTQNSLKPFLTEDELPG-FIHGLTHGTLT-----PAELLQKGT 1430
 Db 1308 --LNAGNINSQSTASSQNTQSSSTYLDRAAGIVITGKEGVLAQAQAKDINIAGQIS 1365
 QY 1431 HQMKGSKLTSVDTSANL-----RAGINLEDGSRKNGVTVARVAGLSASA----- 1479
 Db 1366 NQSEOG-QTRLOAGRDINDLTQVTSKHQATHFDADNHVIRGSTNEVSGSIQTKGDVTL 1424
 QY 1480 --NLAAGSRERSTTSGQSTTASNNRPTFLNGVAGANLTAALGVAHSSTHEKPVGI 1537
 Db 1425 GNLLNAXAEVSSANGTL--AVSAKND---IN-ISAGINTT---HVDASKHTGRSGG- 1473
 QY 1538 PPAFTSTNVSAALADNRTSISLELKRAPVTSNDISELTSTL-----GKHF 1587
 Db 1474 ---GNKLVIITDKAOSHHEAQSTSTEGQVVLQAGNDANILGNSVINDSGTOIQAGNEVR 1530
 QY 1588 DSATTKMLAALKELDDAKPAQLHLQHFSAKQVGVGERYEAVNKLKVIROQAASH 1647
 Db 1531 IGTT-----QTSQSEYTHQTKSGMLMSAGIG-----FTI---GSKTN 1565
 QY 1648 SMELGSASHSTTYNNLSINNDGIVELHKEHFDALPASSAKRLGEMM-----NNDPAL 1702
 Db 1566 TOENQSQSNEHTGSTVGLKGTIT--VAGKHYE-----QIGSTVSSPEGNTIYQA 1615
 QY 1703 DIIKQLQSTPPSSASVSMELKDLRBOATEKAILDGKVGREVG-----LFQDRNNLRV 1756
 Db 1616 SIDIQAAHKLNSNTTQTYEQLGLTVAFSPVTD--LAQAIAVAQSSKQVQSKND-RV 1672
 QY 1757 KSVSVSOS--VSKSEGNTPALLGTSNSAAMSERNIGTINFKYGQDNTPRFT-LEG 1813
 Db 1673 NMAAANAGWQAYQYTGSAONLANGTNAQOVS-----ISITYGQON--ROTTQVQA 1723
 QY 1814 GIAQANPQVASALTDLKKKEGLEMS 1838
 Db 1724 NQAQASQIQAGKTKTLIATGAAEQS 1748

RESULT 6

US-10-282-122A-43811
 ; Sequence 43811, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haeelbeck, Robert
 ; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carz, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA 034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 43811
 ; LENGTH: 6713
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus aureus
 ; US-10-282-122A-43811

Query Match 2.9%; Score 273.5; DB 12; Length 6713;
 Best Local Similarity 17.8%; Pred. No. 6,7e-09;
 Matches 407; Conservative 322; Mismatches 904; Indels 651; Gaps 95;

QY 11 KAAVHTAAHNPVGHVAL-----QQSSSSSPQNAASLAAGKNGKMPRIHQPSA 63
 Db 556 KAAVENALSQVTNAKAGLNGHNLQAKSNANTTINGLQHLTTAQQDKLKQ-QVQQAQNV 614
 QY 64 ADGISAHQKKSFLRGCLGTKEFSRSPAQOQPGTTHSKGATLRDLILARDGTOHEAA 123
 Db 615 A-GVDTV--KSSANTLNGAMGTLNSI-----QNTATKNGQNYLD--ATERNKINYNA 664
 QY 124 APDAARITRGGVKKRRNDDMAGRPVKGSGEDKVPDQ--QKRHQL---NNFGQMROTM 178
 Db 665 VDSANGVINA--TSNPND-----ANAINQIATQVTTSTKNALDGTNHLQAKQTA 712
 QY 179 LSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSSTKATTAHADRV-----EIAQ 231
 Db 713 TNAIDGATNLNKAQK-----DALKAQVTSAQRVANVTSIQQTAN 751
 QY 232 E-----DDSEFOQLHQORLAREBNPPQPPKLGVAFTPIISARFQPKLT--AVA 277
 Db 752 ELNTAMQLOHGIDDENATKQTKYRDAEQSKTAYDOAAKAILINKOTGNSDKAAV 811
 QY 278 ESVLEGTDTTQSPKPSMLKSGAGVTPPLAVLDKGLQALAPDNPALNTLLKQTLGKD 337
 Db 812 DRALQQVTSKDALNGDAKLAEAKAARONLGLN---HITNAQRTALGQINQATTVD 867
 QY 338 TQHYLAHASS-DGSOHLL-----LDNKGHLFDIKSTATSYSLVHNSHPGETKKG 386
 Db 868 GNVTVKTNANTLDGAMNSLQGAINDKATLRNQNYLDADESKRNAYTQAVTAAEG-LINK 926
 QY 387 LAQAGTGSVSDGSKGKISLGSQTSHNKTMLSQPGEAHSL--LTGIWQHPAGARPQG 444

Db 927 QTGGNTSKADVNALNAVTRAKAALNAENLNAKTSATNTINGLNLNLOKDLKHQV 986
 QY 445 E-----SIRLHDDKTHILHPELVWQS-----ADKDTSHLSROA 479
 Db 987 EQAQNUGVGVKDKGNTLNTAMGALRTSTONDNITKTSQNYLDASDSKNNTYAVNNA 1046
 QY 480 DGKLYALKD-----NRT-----LQNLSDNKSSEKLVDKI-KSYSDVQDG 517
 Db 1047 NGVINATNPMDANAINDMANOVNTTKAALNCAQMLAQAKTN--ATNTINNAQDLNQK 1104
 QY 518 QVAILDTDEGRHKMSIMPSLDASPEHSISLSLHFA-DAHQGLLHGKSELEPAQSAVLSHR 576
 Db 1105 KDALKTOVNAQSVS-----DANNVQHTATELANGAMTALKAATADKERTKAS-----GN 1153
 QY 577 LVVADSE-----GKLFSAALPKQDGNELKK 603
 Db 1154 YVNADQEKQAYDSKVNTAENIINGTPNATLVNDVNSAASQVNAKALTALNGDN--LR 1210
 QY 604 AMPQHALDEHFGHDHQSIFPDHDDHGLNALVKNFRQOHAC----- 645
 Db 1211 VAKEHA-----NNTIDGL-----AQLNVQKAKLKEQVQSATTLDDGVQTVKNSSTLN 1258
 QY 646 -----PLGDHQPFGWNLTD-----ALVIDNQLGLHHTNP--EPHE 680
 Db 1259 TAMKGLRDSIANEATIKAGQNYTDASPNRNEYDSAVTAAKAIINOT--SNPTWEPNT 1314
 QY 681 ILMGHLGSLAQEGKLYHFDQLTKGWTAESDCKQ-----LKKGLDGAAYL-- 727
 Db 1315 ITOA--TSQVTTKEHALNGAQLAQAQAKTAKNNLNITSINNAQKDALTRNIDGATTVAG 1372
 QY 728 LKDGVEKRLINOSTSIIKHGTENVFSLPHVRNKPPEGDALQGLINKDDKAQAMAVIGVK 787
 Db 1373 VNOETAKATELNNAMHSLONG-----INDETQTKQKYLDAPSXSAYDQAVNA 1423
 QY 788 YLALTEKGDIRSFOIKPCTQOLEPAOTLSREGISGEIKDIHVDHKNLYALTHEGEVFFH 847
 Db 1424 AKAILTKASQGVNDKAAVEQALQNVNSTKALNGDAKLNEAKAAKQTLGLTLH--INN 1480
 QY 848 QPREAWQ-----GAESSWHKL-----ALPQSSKLSLDMSEHKEPIAFFEDGQ 894
 Db 1481 AQRNALDNEIPOATNVEGVNTVKAQAQGLDGMGOLETSIRDKDTLLQSNYQDADAKR 1540
 QY 895 HQLK-----AGWHAYAAPERGPLAV-----GTSQSQTVFNRLMQGVKVI 936
 Db 1541 TAYQAVNVAATILINKTAGGNTPKADVERAMQAVTQANTALNGIQNL-ERAKQAANTAIT 1599
 QY 937 PGSGLTIVK-----LSAQ-----TCGMTGAG-----RKVSSK----- 963
 Db 1600 NASDLNTKQKALKAQVTSAGRVSAANGVEHTATELNTAMTALKRAIADKADTKASGNV 1659
 QY 964 -----FGERIRAYAF-----NPTMSTPRPIKNA-----YATQ 991
 Db 1660 NADANKROAYDEKVTAAHBHIVSGTPTPL-TPSDVNTAATQVTVNAKTQNGHNLVAKQ 1718
 QY 992 HGWQREGKPLIYEQGALIKQLDAHNVNRHNAPOFDLQSKLETLDLGEHGAELND-MKR 1050
 Db 1719 NANTALDGLTSLNGPQKAKLX-----QVQATTLPNVQTVRD-----NAQTLNTAMK 1767
 QY 1051 FRDELEQSATSVTVLQHQGVGLKNGEINSEFPKSPGKALVQSFNVNRSQGLSKS-LQ 1109
 Db 1768 LRDSIANEATIK-----GQNYTDASQNKQNDYNNAVTAAKAI-----GQTTSPSMIA 1816
 QY 1110 QAVHATPPSAESKLSQMLGHFVSGVDMSHQGEIPLGRDPNDKDTALTSRLILDTVT 1169
 Db 1817 QEINAKQVTRAKQALNG-----QENLRTAQTNAKOHLN----- 1851
 QY 1170 IGELEHA-----DKAKLVSDHDPDADQIKQLRQOPD-----TLREKRYESNPVK--H 1215
 Db 1852 --GLSDLTNAQKDAKROIEGATHVNEVTAQANNADALNTAMTNLKNQIQDQNTIKQGVN 1909
 QY 1216 YTDM-----GFTH-----NKALEAN--YDAVKAFINAFKHEHGVNLTTRTVLES 1258

Db 1910 FTDABEAKKNAYTNVTAQEIILNKAQGTAKOGVETALQNVQRAKNELNGQNVNA-- 1966
 QY 1259 QGSABELAKKLNTLSLSDGESMSFSRSYGGVSTVFVPTLSKKVPVPIPGAGITLDR 1318
 Db 1967 -NAKITAKVALNLTISINNAQKAAKLSQIEGATTVAGVQVS-----TMASE 2012
 QY 1319 YNLSFRTSGGLN-----VSFGDGGVSGNIMVATGHDVMPYMTGKTSAGNAS 1367
 Db 2013 LNTAMSNLQGINDEAATKAAQKYTEADRKOTAYNDVATAAKTLDDKTAGSDNDKNVAVE 2072
 QY 1368 DML-----SAKHISPLDRIGAAVSGTLQ-----GTQNSLKFLEDELPFTHLTHGT 1418
 Db 2073 QALQVNTAKTALNGDARLINEAKNTAKQQLATWSHLTNAQKANLTSQIERGTTVAGVQGI 2132
 QY 1419 LTPAEELLQGIHQMKQSGKLTFSVDT-----ANLDDLKAGIN-----LNEDG 1461
 Db 2133 QANAGTLNQA--NQLRQSIASKDTSSEDIQDANADLQNAVNDVATNAEGIISATNPPE 2191
 QY 1462 SKPNGVTARVS-----AGLSASANLAAGRSRSTTSQGFSTTSASNNRPTFLNG-VGA 1514
 Db 2192 MNPDTTINQKASQVNSAKSALNGEKLAA--VKQTAKSDIGRLTDLNNAQRTAANAEDVQ 2248
 QY 1515 GANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRSTQSISL-----ELKRAEP 1569
 Db 2249 APNLAAYTAAKNKATSLNTAMG-----NLKHALAEKDKNTKRSVNYTDADQPKQAYD 2300
 QY 1570 VTSNDISELTSLTGKHKFQKSAATTKMLAALKBELDDAKPAEQHLHQHFSKADVVGDER-- 1627
 Db 2301 TAVTQAEAITNANGSNAE--TQVOAALNQLNAQ-----NDLNGDNKVA 2343
 QY 1628 -----YEAVNLKLVIROQ-----AADSHSMELGSA----- 1654
 Db 2344 QAKETAKRALASYNLNNAOSTAATSOIDNATTVADVTAQNTANELNTAMGOLQNGIND 2403
 QY 1655 -----SHSTTYNNLSRINNDGIVE-----LLHKHFDPAAL----- 1683
 Db 2404 QNTVKQOVNFTDADQKKDAYTN-AVTNAQOILDKANGQNNMTRKAQVEAALNQVTTAKNAL 2462
 QY 1684 -----PASSAK-RLGEMNN-NDPALKDIILQOSTPFPSSASVSVMELK-----DGLREQ 1729
 Db 2463 NGDANVRQAQSDAKANLGLTTLNNAQKQDLTSQIEGATTVNGVNSVTKRAQDLDDGAMQR 2522
 QY 1730 TEKAILDGKVBREVGVLFDORNNLRVKSVSQS-----VSKSEGFN-----TPALLIGT 1780
 Db 2523 LESAIANKOQTKASENVIDADPTKKTAFDPAITQAESYLNKDHGNTKDKQAVEQAIQSVT 2582
 QY 1781 SNSASGMERNIGTINFKYGO-----DONTPRFTLEGGIAQANPOVASALTDLKKBEGL 1834
 Db 2583 STENALNGDANLQCAKTEATQADLNLQTLNTPQKTLKQVNA--QVSGVTDLKNSAT 2640
 QY 1835 EMKS 1838
 Db 2641 SLNN 2644

RESULT 7
 US-10-311-879-28
 ; Sequence 28, Application US/10311879
 ; Publication No. US20030186275A1
 ; GENERAL INFORMATION:
 ; APPLICANT: University of Sheffield
 ; TITLE OF INVENTION: Antigenic Peptides
 ; FILE REFERENCE: toxin
 ; CURRENT APPLICATION NUMBER: US/10/311,879
 ; CURRENT FILING DATE: 2003-03-17
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 2659
 ; TYPE: PRF
 ; ORGANISM: Staphylococcus aureus
 US-10-311-879-28

Query Match		2.8%; Score 267.5; DB 14; Length 2659;
Best Local Similarity		18.2%; Pred. No. 3.8e-09;
Matches 374; Conservative 298; Mismatches 780; Indels 599; Gaps 86;		
QY	8 TEHKAHVTAHNPVGHGVALQ--GSSSSSPQNAASIALAEGKNGKMPRIHQPSAAD 65	1766 AQRQLSQINGAHQIDAVENTIKQ-----ATNLN SAMGNLRQAVADKQVKRT 1814
DB	933 TOYNAMENAOQOINTAKTEAOQVINNERATPQQVSDALT-----KVRAAQTQID 982	960 -----VSSSESEIRAYAFNPTMTSPRIKNAAYATQHQGREGKLP 1002
QY	66 GISAAHQKPSLRCGLTKKFSRPAQOGPGTTHSKGATLRDLILDGGETOHEAAAP 125	1815 EDYADADTAKONAYNSAVSS--AETIINQTTNPTMSVD--DVRNATSAVTSNKNALNGYBK 1871
DB	983 QAKALLQNKEDNSQ--LVTSKNNLQSSVNOVPST--AGMTQOSIDNNAKREAEITEIT 1037	1003 LYEMOGALIKOLIDAHNVRHNAPOPLQSKLETLD-----LGEHGAEL-----L 1045
QY	126 DAARLTRGGVKKRRMDD-----MAGRPVWKGSGEDKVPTQOKRHQLNFMQCMQT 177	1872 LAQSKTDAARALDALPHLNNAKADVKSKINAASNIAGVNTVKQOQGTDLNTANGNLQAI 1931
DB	1038 AARQVIDNGDATAQQIISDEKHRVDNALTALNOAKHDLTADTHALRQAVQQLNRTG---T 1093	1046 ND-----MKRFRDEQSATRSVTVLQHQGVLL--KSMGEINSEFKPSGKALVQSFNN 1098
QY	178 MLSKMAHPASANAGR-----LQHSPPHIPGSHHEIKERPVGSTSKATTA--HADREVEIAQ 231	1932 NDEQTLLNSQNVQDATPSKKTAYTNAVQAAKIDLNKSNQNKTKQOVTEAMNQVNSAKNN 1991
DB	1094 TTGK--KPASITAYNNSSTRALQSDLTSAKSNANALIQPIRTVQEVQSALTNNRV----- 1147	1099 RSGQDLSKLOQAVHATPPSAESKLOSMGLHFVSAGVDMSHOKGEBIPGRQDPNDKXTAL 1158
QY	232 EDDSEFQOLHQ-----QLRARERENPPQPKLGVATPISARFQPKLTAV----- 276	1992 LDGTRLLDQAK-----TAKQQLNNMT--HLTTA-----QKTNL 2023
DB	1148 --NERLTQAINQLVPLADNSALKTKTKLDBEINKSVTTDGMTQSSIOAYENAKRAGQTE 1205	1159 TKSRLLDVTITIGELHELADKAKLVSDHKPDADQIKQLRQOQFDTLREKRYESNPVKHYD 1218
QY	277 ---AESVLEGTDTTQSPKPO-----SMLKSGAGVTPPLAVTLDKGLQLAPD----- 321	2024 TNO--INSQTTVAGVQTVQSNV-----NTLDQAMNTLURQSIANKDATKASED 2068
DB	1206 STNAQNVINNGDATDQQAABKTKVEEYKNSLQKAIAGLTPLDPLAQTAQTQLQNDIDQP 1265	1219 MGFTHNKALEANYDAVKAFINAFKKEHGVNLTRTVLESQGSABELAKKLXNTLLSLDSG 1278
QY	322 -----NPPALNTLLK--QTLGKDQHVLAHH-----ASSDGSQH 353	2069 Y-----VDANNDKQATAYNNAVAAAEIINANSNPMNPSTITQKAEQVNSSKTALNGD 2121
DB	1266 TSTGTMTSASIAAFNEKLSAARTKIQEIDRVLASHPDVATIRQNVTAANAAKSALDQARN 1325	1279 BSMFSRSYGGGVSTVFPVPTLSKKVPVPIPGAGITLDRAYNL-----SPRTSGGLNVS 1333
QY	354 LLLDNKGLPDIKS-----TATSYSVLHNSHPGEIKGLAQAGTGSVVDGKSGKISL 406	2122 ENLAAAKQN-----AKTYLNTLT-----SITDAQKNLISQITSATRVSGVDTVK 2166
DB	1326 GLTVDKAPLENAKQLOYSIDTSTTGTQDSINAYNAKLTARNKIQOIN-----QVLA 1381	1334 -----FGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPLDLRIG 1383
QY	407 GSGTQSHNKTMLOPGEAHRSLTGIWOHPAGARPQGESRLDHDKTHIHLPELGVWQS 466	2167 QNAOHLDOAMASLONGINNESQVSKSEKRYRADTNKQOEDYNAI--TAAKAILN----- 2218
DB	1382 GSPTVEQINTNTSTANQAKSLD-----DHARQALTP-----DKAPLOTAKTQLRQS 1427	1384 AAVSGTLQGLTQNSLXFKPL-----TEDELPG-----FIHGLTHGTITPPABLL 1425
QY	467 ADKDT-----HSQLSROADGKL-----YALKDNRTLQNLSD-----NKSSEKLIV 505	2219 ---KSTGPNTAQNAVEAALQRVNNAKDALNGDAXLIAAQAQAQHLGLTLH--ITTAQ-- 2271
DB	1428 INQFTDTGTTASINAYNQKLOARQKLTEINQVNGNPTVQINQKVTBANQAKOQLN 1487	1426 QKGTEHQMKQSKL--TFSDVDTSAN-LD-----LRAGIN-----INEDGSKPNG 1466
QY	506 DKIKSYVDQGGQVAILTDPGRHKWSIMPDSAPESHISLSLHFADAHQGLLHGKSEL 565	2272 RNDLTNQISQATNLAGVESVKQNAISLDGAMGNLQTAINDKSGTLASQNFILDADQEKNA 2331
DB	1488 TARQGLTLDRQPALTTL-----HGAS---NLNQAQQNNFTQQINAAQNAHALETIKSI 1538	1467 VTARVSAGLSASANAAGSRERSTTSQFGSTTSASNNRPTFLNGVGAGANLTAALGVAH 1526
QY	566 EAQSVAIHSHGLVVDSEGLFSAAI PKQGDGNELKMKAMPQHALDEHFGHDHIOISGFFH 625	2332 YNQAVSAA--ETILNKQTPNTAKTAVEQ---ALMNVNNAKHALNGT---QNLNNAKQAA-- 2383
DB	1539 TALNTAMTKLKDSVAD-----ASVSKTKDALDQOQNLQRAK-----TEATNAI 1630	1527 SSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTS-----NDISELST 1581
QY	626 DDHQNLALVKNRPQOCHACPLGNDHOPGWNLTDAVIDNQJLGHHTWPEPHEILDMG 685	2384 -----ITAINGASDLNQKQKDALKAQANGAQRVSNQADQVQHNATELNTA 2427
DB	1555 -----NNTIKSDQNYTDATP--ANKQAVDNVANAAGVIGET-----TNDVN 1597	1582 LG--KH-----FKDSATTQMLAALKELDDAK-----PASQLHLIQ 1614
QY	686 HLGSLALQEGKLHYFDQLTKGWTGAESDCQKGLKGLDGAAYLLKDGVEKRLNINQSTSSI 745	2428 MGLTKHAIADKNTLASSKYVNADSTKQNAVTTKVTNAEHIISGTPVTVPSEVTAAN 2487
DB	1598 TVNQKA-----ASVSKTKDALDQOQNLQRAK-----TEATNAI 1630	1615 QHFSAK--DVVGDERY--EAVRNL-----KKLVIRQQAADSHSMELGSAHSTT 1659
QY	746 KHGTENVFSLPHVNKPEPGDALQGLNKDKQAQAMVIGVKNYLYALTEKDIFRSQIKPG 805	2488 QVNSAKQEIINGDELRERAKQKNANTAI DALTQNLTPQAKUKEQVQANRLB-----DVQTV 2543
DB	1631 THASD-----LNQAQKNALTOQVNSAQNVQA-----VN-----DIKQT 1663	1660 YNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNDPALDKI IKLOSTPFSSASVS 1719
QY	806 TOQLERPAOTLSREGISGELKD IHVD-----HKQNLV--ALTHEGEVHPQPREAWQNG 856	2544 QTNGQALNN-----AMKGLRDSIANETTVKTSQNTYTDASPNNQSTYN 2585
DB	1664 TQSL--NTAMTGLKRGVANHNOVQSDNYVNAATNKKNDYNNAYNHANDII-----NG 1714	1720 MELKD--GLREQTEKAILDGKVBEEVGLFPQDRNNLR--VKSVSVSQSVSKSEGFNTPAL 1776
QY	857 AESSWHKLALPQSESKSLDMSHEHKPIATFEDGSGHOLKAGGWAYAA----- 907	2586 SAVSNARKIINQTNPTMDTSITQATTOVNNNAKNGLNGAENLRNAQNTAK--QNLNTLSH 2644
DB	1715 --NAQHPVITPDSVNNALSNVTSKEHAL-----NGEAKLNAAKQANTALGHNLNLN 1765	1777 LLGTSNSAAMS 1787
QY	908 PERGPLAVGTSGSQTV--FNRLMQGVKGVIPGSGGLTVKLSAQTGMTGAEGRK----- 959	2645 LTNNQKSAISS 2655

US-09-815-242-12996
 ; Sequence 12996, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITFA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12996
 ; LENGTH: 6281
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus aureus
 US-09-815-242-12996

Query Match 2.8%; Score 262; DB 9; Length 6281;
 Best Local Similarity 17.9%; Pred. No. 4.1e-08;
 Matches 401; Conservative 319; Mismatches 856; Indels 668; Gaps 94;
 QY 27 ALQGGSSSSPONAASLAARGKNR-----GKMPRIHOPSTAADGISAHQKK 75
 Db 408 SIQNTATKQGN---YLDATERKNTYNNVAVSANGVINATSPNMDANALQIAQVT 464
 QY 76 SFLRGLGCTKKFGRSAPQGPQTHSKGATLRDLARDGETQHEAAAADARLTRSGG 135
 Db 465 S-----TKNALDGTTHLNTQAKQTATNAI-----DGATNLNKAQKADALKAQVTS 508
 QY 136 VKRNMDDMAGRPVKGSGEDKVTQKRRHLNFGQMRQTLMSKWAHPASANAGRLQ 195
 Db 509 QRVANVT-----IQOTANLNTAMGLOQHGIDDENATKQTKYRQAEQSKTAYDAQV 562
 QY 196 HSPHPHSGHREKEEVPVGSSTKATTAHADVIAQEDDDSEFQOL-----HQORFLA 247
 Db 563 AAAKAI-----LNKQTGNSDKAAV-----DEALQVTSKDALNGAKIA 603
 QY 248 RERENPPQPKLGWATPISARFQPKLFAVESVLEGTTTQSPKQSMKLGSGAGVTP 307
 Db 604 EAKAAKQ--NLGTNLHI-----TNAQRTDLEGGQINQATTVDGVNTVKTIN----- 646
 QY 308 AVTLDKGLQLAPDPPALNTLLKQTLGKD---TOHVL-AHHASSDGSQHLLDNKGHL 362
 Db 647 ANTLD-----GAMNSLQGSINDKATLNRQNLVDADSKRNAYTQAVTAEGIL 695
 QY 363 FDIKSTATSYSLHNSHPGEIKGLAQAGTGSV-----SVDGSKGKISLGSQTOSH 413
 Db 696 NKQTGNTSKADVNDALNVAIFAKAALNAGDNLNRAKTSATNTIDG-----LPNLTQLQ 749
 QY 414 NKTMLSQCEAHRSLLTGWIQHPAGARPOGESIRLHDKTHILHPELGVWQS----- 466

Db 750 KDNLKHQVEQA-----QNVAG-----VNGVKDKGNTLNTAMGALRTSIQNDNT 792
 QY 467 -----ADKTHSQRQADGKLYALKD-----NRT-----LQ 493
 Db 793 TKTSONYLDASDNKNYNTAVNNANGVINATNPNMDANANGMANQVNTTKAALNGAQ 852
 QY 494 NLSDNKSSEKLVDKI--KSYSVQDQGVVAITLTPGRHKMSIMPSLDASPESHLSLHFA 552
 Db 853 NLAQAQTN--ATNTINNAHDLNQKQKDALKTQVNNARVS-----DANNVQHTATELNSA 905
 QY 553 -DAHQGLHKGSELEASVAISHGRIVADSE-----583
 Db 906 MTALKAAIAADKERTKAS-----GNVNADQEKRAYDSKVTNAENLISCTPNATLTVND 959
 QY 584 -----GKLFSAIIPKQDGNELMKAMPQHALDEHFHGHQISGFFHDDHGQNAL----- 634
 Db 960 VNSAASOVNAAKTALNGDNN--LRVAKHA-----NNTIDGLAQLNNAQAKLKEQV 1009
 QY 635 -----VKQNFQOHACPLG-----NDHQPHPGWNLTD-----A 662
 Db 1010 QSATTLTGVTQVKNSSQTLNTAMKGLRDSIANEATIKAGQNYTDASPNRNEYDSAVTAA 1069
 QY 663 LVIDNQLGLHHTNP--EPHEILDMGHLSLALQEGKLHYDPDLTKGTWGAESDCKQ----- 716
 Db 1070 KAIINOT-----SNPTMEPNTITOV--TSQVTTKEALNGARNLAQAKTIAKNLNLTSL 1123
 QY 717 -----LKKGLDGAAYL--LKDGEVKRLNINOSTSIIKHGTENVFSLPHVRNKPPEGDA 767
 Db 1124 NNAQKDALTRS IDGATTVAGVNETAKATELNAMHSLQNG-----INDETQPKOT 1174
 QY 768 LOGINKDDKAQAMAVIGVKNYLALTEKGDIRSFIKPGCTOOLERPACTLSREGISGLKD 827
 Db 1175 QKYLDAEPSSKSAIDQAVNAAKAILTKASQNVDKAAVEGALQNVNSTKALNGDAKLINE 1234
 QY 828 IHVDHKNLYALTHEGEVPHQPREAWON-----GAESSWHKIL-----ALPQSESKL 874
 Db 1235 AKAAKQTLGLTH---INNAQRTALDNEITQATNVEGVTNKAQAQQLDGAQWQLETSI 1291
 QY 875 KSLDMSHEHKPIATTFEDGSQHOLK-----AGWHAYAAAPERPLAV-----G 916
 Db 1292 RDKDTTLQSQNYQDADDAKRTAYSAQVNAATAILNKTAGNTPKADYERAMQAVTOANTA 1351
 QY 917 TSSQTVFNLMOGVKGVIPGSGLTVK---LSAQ---TGGMGTGASG-----957
 Db 1352 LINGIQNL--DRAKQAANTAITNASDLNTKQKALKAQVTSAGRVSAAANGVHTATELNTAM 1410
 QY 958 ---RKVSSK-----FSEIRAVAFNPTMTSTPRP-----IKNAA- 987
 Db 1411 TALKRATADKAAETKASGNVYNADANKQAYDEKVT--AENIVSGTPTTTLTPADVNTAAT 1469
 QY 988 -----YATQHGMOGREGKPLXYEMQGALIKQLDAHVNRHNAPOPDLOSK 1031
 Db 1470 QVTNAKTQLNGNHNLEVAQNANTAIIDGLATSLNGPQKAKLKE---QVQATTLTPNVQTV 1525
 QY 1032 LETLIDLEHGAELLND--MKQPRDELEQSATRSVTVLQHGQVLKSGNINSEEPKPCA 1090
 Db 1526 RD-----NAQTLNTAMKGLRDSIANEATIKA---GQNYTDSQNKQTDYNSAVTAKA 1575
 QY 1091 LVQSFNVNRSQDLSKSLQ-QAVHATPPSABSCLKQSMGLGHEFVSAGVDMSHQGEIPLGRQ 1149
 Db 1576 II-----GQTTSPSMNAQEIQAQKQVTAQKQALNG-----1606
 QY 1150 RDPNDKTLTKSRILDTVTIGELHELADKAKLVSDHKPDA-----DQIKQLRQ 1198
 Db 1607 -QENLRTAQTNWAK-----QHLNGLSD---LTDQAKDAVKRQIEGATHVNEVTOAQN 1653
 QY 1199 QFD-----TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFLNAPKKEHGVNLT 1251
 Db 1654 NADALNTAMTILKNGIQDQNTIK---QGVNFTDADAKRN---AYTNVATQAEQILNKA 1706
 QY 1252 -----TRTVLESQGSAAE-----LAKKLNLTLLSDSGESMSFSRYSY 1288
 Db 1707 QGPNTSKDGVETALENVQRAKLNELNGONVANAKTAKNALNLTSLINNAQKALKSQIE 1766

Db 1044 VRDYHETIYKERIENRPAHITVGNLTASQGNLWLNKDSRI-----VVGGRIITDLDNQ 1097
 Qy 582 -----SEGLFSAIPKQGD-----GNELKWKAMPQHALDEHFCHDHOISGEFFH 625
 Db 1098 KEITNOSTTGKRTDAVGQWDSVTKKGYSGRRQRREKNT-----SYHDTQL--FTH 1151
 Qy 626 DRHQNLALVKN-----NFRQOHAC-----PLGNDHQFHPG--WNLTDALVIDNQLG 670
 Db 1152 DFPTPVSIVQOAAASFPQAAAKLIDGASTAAVNGQRIHTGNVSLNNAATVTLNPS 1211
 Qy 671 LHNTNPEPHEILDMHGLSIALQEGKHYDPQITKGTWGAESCKQKGLDGAAYLLKD 730
 Db 1212 LYTHPD-----NKGW-----LIETDPQFADY--RRWLGSIDYMLQQLKLTNHLHLRGLD 1259
 Qy 731 GEVKRLNINOSTSSIKHGTENVFSLPHVRNKPBGDALQCLNKKDDKAQAMAVGVNKKLA 790
 Db 1260 GYTEQKLVN-----EQIHQLTGYRR-----LDGYKSDDE-----QFKA 1292
 Qy 791 LTEKG--DIRSFOIKPG--TOQ-----LERPAQTLREGISGELKDIDVHDHK 834
 Db 1293 LMDNGLTAAKTFTGLTGLIALSAEQVARLTSIVVMENQTVTL-----DGSTQTVLV--P 1345
 Qy 835 NLVALTHEGVFQPREAWNGAESSWHKALIPQESKLSLSDMSHEHKPIATFEDGSG 894
 Db 1346 KVALARKGDL-----NTSGGLISAEQVLLKQNGNLNLS--GTIA-----GRQ 1387
 Qy 895 HOL-----KAGWHAYAAPERGFLAVGTSGSTVFNRMLQGVKVPICSGGLTVKLSA 947
 Db 1388 AVLIQARNINSNGNIQADQILGKAEKSINIDGGVQVQAGRL-----TAQAQINLNGTT 1441
 Qy 948 QTGG-----MTGABGRKSVSKFSERIRA-YAFNPTWTSTPRPKNAAY-- 988
 Db 1442 QTSNERNNGTATDRMAGINVGVSHTSQVDNRNRTSDGILSLHAGNDINLNAATVSNQVKG 1501
 Qy 989 ATQGWGREGRLKPLYEMQAGALIKQDAHVNRHNAPOPDLOSLETLDLGEHGAELL--N 1046
 Db 1502 TQITAGNNILNGTIHTHEAREAYGLDDENHRHVROSTEVGSSIRT-----QNGAMLRAGN 1557
 Qy 1047 DMKRFRDELCOSATRSVTVLGOHQVL-----KSNGEINSEFKPSP----- 1087
 Db 1558 DLKTRQELADEGKTVLAGROVNISEGRQITELDASVSKSGKILSSYKTHDPYFSSH 1617
 Qy 1088 -----GKALV-----QSFNVN-----RSGDLSKSLQOAVHATPPSAESK 1122
 Db 1618 DEAVGNIQGGKMIVAAGQDINVRGSLNLSKGTVLKAGNDIDISTAHNRYTGNEXYESK 1677
 Qy 1123 LQSLMGLHFVAGVDMHQKEIPLGRQR--DPNDXTALTCSRLL-----DVTIT----- 1170
 Db 1678 KSGVMG--TGGLGFT-----IGNRKTDDTDRTNIVHTGSIIGSLNGDVTTVAGNRY 1727
 Qy 1171 ---GELHELADKAKLVSDHPDADQIKLQROQFDTLREKRYESNPKHYTDMGFT---HN 1224
 Db 1728 RQTGSTVSSPEGRNTVTAKSIDVES-----ANNRYATDYVHTREQKGLTVALNV 1776
 Qy 1225 KALEANYDAVKAFINAFKEHGVNLTTRTVLESQGSAAELAKLKNLTLSDSGESMSFS 1284
 Db 1777 PVVQAAQNFVQAAQNVGSKNKNRVNMAAANAQW-QYAAQQAQMOQFAPSSAGQGN 1835
 Qy 1285 RSYGGGVSTVFVPTLSK-----KVPVPVPGAGITLDRAYNLSFSRTSGG---LNV 1332
 Db 1836 QSSGISVSVTYGEOKSNEOKSRVTEAAASQIIGKQT-----TLAATSGSGQSNINI 1888
 Qy 1333 SFGRDGVSNGIMVATGHVDM--PYMTGKTSAGNASDW-LSAKHKISPDLRIGAAVSGT 1389
 Db 1889 TGSVDIGHAGTALIADNHRILQSAQDQSGSEKSKSGWAGVAVKINGIRFGITAGN 1948
 Qy 1390 L-QGTQLNSLKFUKTEDELPFGIHLTH-----GTLPAELLQKIEHQM 1433
 Db 1949 IGKKEQGG-----NTTHRHVHTAGOTTIRSGDITTLKGAQLIGIQIADT 1997
 Qy 1434 K-----QSKLTFSDVTSANLDRAGINLNEDGSKENGVTARVSAGLSASANLAAGS- 1485
 Db 1998 RNLHIESVDPTQSKQOQNGNVQTVGVGFSAGSYRQSKVKADHASVTCQSGIYAGED 2057

Qy 1486 -----RERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHS-----STHEGKPVG 1536
 Db 2058 GYQIKVRDNLTKGGIITSSQ-----SAEDKGNLFTQATLTHSDIQHNSRYEKGKSGF 2110
 Qy 1537 IPPAFSTVNSAALALDNRTSQSISLELKRAEPV-----TSNDISELT-----STLGKHPK 1587
 Db 2111 IGSFPLNG-----GWDGTVTKQGRPTDRISPAAGYSGDGSKNSTTRSGINTNRNIHIT 2165
 Qy 1588 DSA--TTKMLAALKELD-----DAKPAEQHLI--LQOHFSAKDVVGDREYEAVERNKK 1636
 Db 2166 DEAGQLARTGRTAKETEARLYTICIDTETADQ--HTGRKNSFD-KDAVAKE-----INLQR 2218
 Qy 1637 LVIRQ-----QAADSHSMELGSAHSSTYNNISRNNDGIVLELLHKHFDPAALPASSAKR 1690
 Db 2219 EVTKFGRNAAQAAVAADKLGNTQSVRYEQEARTLLEDELQNTDSEAKAAIRAS----- 2274
 Qy 1691 LGEM 1694
 Db 2275 LGQV 2278

RESULT 10

US-10-282-122A-70580
 ; Sequence 70580, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 70580
 ; LENGTH: 6641
 ; TYPE: PRN
 ; ORGANISM: Staphylococcus epidermidis
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (6636)..(6636)

OTHER INFORMATION: X-any amino acid
US-10-282-122A-70580

Query Match 2.7%; Score 258.5; DB 12; Length 6641;
Best Local Similarity 17.7%; Pred. No. 8.2e-08;
Matches 396; Conservative 319; Mismatches 805; Indels 723; Gaps 90;

QY	33	SSSSQNAASL-AAGKQKGRMPRIHQPSSTAADGISAAHQKSFSLRGCLGKTKFSSRS	91
DB	2694	ATQSVQAEALHGAELNQDK---QTSSTELDGLTDLTDAQRE-KLREQINTNSRDD	2748
QY	92	APQ--GQPCTTHSKGATRDLLARDG-----ETQHEAAAADAAR	129
DB	2749	IKQIEQAKALNDAMKKLEQVAQKDGVHANSYDYNEDSAQDAYNNALKQAEDIINSS	2808
QY	130	-----LTRSGGVKRRNMDMAGRPVVKGGSGEDKVPTQKRRHQLNLFNGQWROTMLSK	181
DB	2809	NPFLNAQDITNALNIIKQADNLHGAQLQ---QDKNTTNAIGNLHNLNFPQKDALIQ	2864
QY	182	MAHPASA-----NAGRLQHPHPHPPGSHHEIK	209
DB	2865	AINGATSRDQVAEKLKEAEALDEAMKQLDQVNDQDQISNSPFFINEDSDKQTYNDKIQ	2924
QY	210	--EEPVGSTSKAT-----TAHADRVETIAQEDDDSEFPQQLHQORLARE	249
DB	2925	AKEIINOTSPTLDKQKIADTLQNIKDANNLHGDQ-KLAQSKQDANNQLNHLDDLTEE	2983
QY	250	RENPPQPKLGVAATPISARFQPKLTAVAESVLGDTTQSPKLPQSMKGGAGVTPLAV	309
DB	2984	QKN-----HFXP-----LNNADTRDEVN	3003
QY	310	TLDKGLQALPDNPALNTLLKQTKGTQHYLAHHSAGSQHLLLDNKNKGLHFDIKSTA	369
DB	3004	-----QLEIAKQLNGDMSTLHKVINDKQIQHLSNVINADNKKQNYDN-----AIK	3050
QY	370	TSYSLVHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTSGHNTKMLSPQGEAHRSL	429
DB	3051	EAEDLIH-HPDLDHKAQ-----DLINKIDQAHNEL	3082
QY	430	TGIWQHGAAGAR-----PQGESIR-----LHDDKIHILHPELVGMQS	466
DB	3083	-----NGESRFKQALDNALNIDSLNSLVPRQTVKDNINHVTTLESQAEL	3130
QY	467	ADKUTHSOLSRQADQKLVAKD-----NRTLQNLSDNKSSEKLVDKIKSY	511
DB	3131	-----QKAKELNDAMKMRDSIMNQEQIRKNSNYTINEDLAQQNAYNHAVDKINNII	3183
QY	512	-----SVDRQGOVAILTDPGRHKMSIMPSLD	538
DB	3184	DNAIWDPOIIOKATQDINTAINGLKGDKQKLODAKQDQITNFTGLTEPQKQAL	3238
QY	539	ASPESHISLSLHFADAHQGLHKG---SELEAQSVAISHGRILVAD-----SEGKLF	587
DB	3239	---ENILNQTSRANVAKQLSHAKFLNGKMBELVAVAKASLVQRNSNYINEDVSEKEAY	3295
QY	588	SAALPKQGDGNE-LMKKAMPOHALDEHFDHDIQSGFFHDHQLNALVKNFRQOHACP	646
DB	3296	EQAIKAK---GGEIINSENNPTISSDINRTIQEINDAEONLHGD-----NKLQAQAEI	3345
QY	647	LGNHQHHPGNLTDALVIDNQLGLHHTNPPEHLLDMGHLSLALQKGLHYFDQTKG	706
DB	3346	AKNEIQLNDGNSAQITKLIQDGRITTKPAVTQKLEBAKAINQAMQOLKQSIADK	3401
QY	707	WTGAESDCKQLKKGIDGAAYLLKQGEVKNLN-----QSTSSIKHGTENVFSLPHVRN	760
DB	3402	-----DATLNSNYSINESEKKLAYDNVSAQELINQLNDPTMDISINQAITQ	3450
QY	761	KP-BFGDALQGLNKDDKAQAMAVIGVKNYLAITEKGD-----IRSFQKPGTQOLERP	813
DB	3451	KVIQAKDSLHGAANKLAQNAQADSNLIINGSTNLNDKQKQALNDLINHAQTKQQAIEIT	3510
QY	814	QTLSEGISGLKDIHVD-----HKQNLVALTHEGEVHQPPEAWQNGAESSSW	862

DB	3511	NKLINNE--MGTLKTLVEEQSNVHQSKY-INEDPQVQNIYNDISIQKREILNGTITDDVLN	3567
QY	863	-HKLALPQSESKLSLSDMSHEHKPATPFDGSGH-----QLKAGWHAYAAPERG	911
DB	3568	NNKIADAIQNIHLTKNDLHGDKQKLOKAQQADATNELNLYNLNNSQSQSHDEINSAPSR	3627
QY	912	PLAVGTSGSQTVFNRMLQGVKGVIPGSGLTIVKLSAQTCGTGAEGRKVSXSFSEIRAY	971
DB	3628	EVSNDLNHAKAL-NEAMROLENEVALENSVK-KLSDFIN-----EDEAAQNEYSNALQKA	3680
QY	972	--AFNPTMST-----PRPKNAAYATQGMQGRGLKPLYEMOGALIKQLD---AHN	1018
DB	3681	KDIIINGVFSSTLTKATIEDALLELQNAVESLHGEQKLEAKNQAAVEIDNLQALNPGQVL	3740
QY	1019	-----VRHNAPODLSQKLETLDLGHEGAELNDMKRPFDELEQSASTRSVTVLGHQGV	1073
DB	3741	AKTLVNGQASTPVEQAL-----QKAKELNEAMKALKTEINKK-----EQIK	3783
QY	1074	KNGEINSEFPKSPGKALVQSFN--VNRSGQ-----DLSKSLQQAQVHATPPSAESKL	1123
DB	3784	ADSRVYNAD-----SGLQANYNSALNYGSOIIATTPPELNKDVINRATQIKTAENNL	3837
QY	1124	--QSMGLHFVSAG-VDMGHQKGEIPLGRQRPNDKTALTLSRLILDTVTIGELHELADKA	1180
DB	3838	NGQSKLAELAKSDGNGQSIIEHLQ-----LTQSQK-----DKQHDLLINQA	3875
QY	1181	KLVSDBKPDADQI-----KOLRQOQDFTLREKRVESNPVKHYTD-----MGTFHNKALBA	1229
DB	3876	QT-----KQOVDIIVNNSKQLDNSMMQLOQIIVNNDNTVKNQSDFINEDSSQDAYNHAIQA	3931
QY	1230	NYDAVKAFINAFKKEH-----HGVNLTTR-----	1253
DB	3932	AKDLITAHPTIMDKNQIDQAIENIKQALNDLHGSNKLSEKKEASEQQLNLSLNGQD	3991
QY	1254	TVLESQGGAEI-----AKKLNTLLSI-----DSGESMSPSRSYGGGVSTVFVPT	1298
DB	3992	TILAHIFSAPTRSQVGEKIASAKQLNNTMKALRDSIADNNEITLQSSKYFNEDEQ	4046
QY	1299	LSKKVPVPIPCAGITLDRAYNLSPRTSGGLNVSFGRDGGVSGNIMVATGHDVWPMYTG	1358
DB	4047	-----QNAVQAVNKAANIIN-----DQPTPYMANDEIOSVLNE	4080
QY	1359	KKTSAGN-----ASDWLSAKHKISPDLRGAASVGTLOGTLQNSLK-----	1399
DB	4081	VKQTKDNLHGDKQKLANDKTDAQATILNALYNLQQRGNLETKVQNSNRPPEVQKVQVLAN	4140
QY	1400	-----FKTEDELPGFHLGHTGLTPAELLQKQIEHOMQKQSKLTFSDVTS--ANLD--	1450
DB	4141	QLNDAMKKLDDALTG-----NDAIKQTSNY-INEDTSQQVNFDEY	4179
QY	1451	LRAGINLNEGSKPKNGVTARYSA-----GLSASANLAAGSRERSTTSQGFSGTTS	1500
DB	4180	TDRGNIVAEQTNPNMSPNTINTIADKITEAKNDLHGVLKQKQAOOSINTINQMTGLNQ	4239
QY	1501	ASN-----NRPTFLNGV--GAGANLTAALGVAAHSSTHEGKPVGIF	1538
DB	4240	AQKELQAEIQOQTRSEVHVQINKAQALNDMMTLQSIITDEHEVKQTSVINETVGNQ	4299
QY	1539	PAFTSTNVSAALDNRTSOSI--SLEKRABPVTSN-DISELTSTLGHKFKDSATTKWL	1595
DB	4300	TAYNNA-VDRVKQIINQTSNPTNPNPLEVERA---TSNVKISKALHGERELNDKNKSTP	4355
QY	1596	AA--LKELDDAKPAEOLHILQOHP-----SAKQVVG-----	1624
DB	4356	AVNHLNLAQAKQKALTHEIQATIVSQVNNIYNKAKALANDMKKLDIVAQQDNVRQSN	4415
QY	1625	-----DERVEAVNLKKLVIRQQAAD-SHSM	1649
DB	4416	NYINEDSTPQMYNDTINHAQSIIDQVANTMSHDEITENAINNKHAI---NALDGEHL	4472
QY	1650	ELGSASHSTTYNNLSRIN--NDGIVELLHKHFDAAALPASSAKRLGEMMNDPALKTIK	1706
DB	4473	QAKENANLLINSINDLNAPQROAINELVNE-----AQTRKVAEQLOSAQALNDAMK	4525

QY 1707 QLOSTPSSASVSMELK-----DGLREOTEKAILDGKVGREYGVLFQ-----DRNLRV 1756
Db 4526 HLRNSIQSSVQESKYINASAKQYNHAV-----REVENINEQHPTLDKEILQ 4579
QY 1757 KSVSVQSVSKSGFNTPALLLTSNSAAMS-----MERNIGTFNFKYQDQNTPRPTLE 1812
Db 4580 LTDGVNQA-----NNDLNGVELLDADKQNAHQSIPTLMHLNQAQNALNEKINNVAVTRTEVA 4636
QY 1813 GGIAQANPQVASALTDLKKEGLE 1835
Db 4637 AIIQOAK-----LLDHAMENLE 4653

RESULT 11

US-10-661-809-23
; Sequence 23, Application US/10661809
; Publication No. US20040101919A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR
; FILE REFERENCE: P07741US01/EAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 10203
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-10-661-809-23

Query Match 2.7%; Score 258.5; DB 16; Length 10203;
Best Local Similarity 17.7%; Pred. No. 1.7e-07;
Matches 396; Conservative 319; Mismatches 805; Indels 723; Gaps 90;
QY 33 SSSPQNAASL--AAGKNGKMPRIHPSTAAAGISAAHQOKSFSLGCLGPKKFSRS 91
Db 6262 ATQSVQNAQALHGAELNODK-----QTSSELDTLDTLDAQRE-KLREQINTNSRDD 6316
QY 92 APQ--GPGTTHSGATLRDLARDG-----ETQHEAAPDAAR----- 129
Db 6317 IKQIEQAKALNDAMKLEQVQAKQGVHANSDYTNEDSAQKDAYNALKQAEIDINSS 6376
QY 130 -----LTRSGGVKRNMDMAGRPMVKGSGGKEDVPTQOKRHQLNMFQMKQTLMSK 181
Db 6377 NPENLAQDITNALNNIKQAQNLHGAQKLQ-----QDKNTTNAIGNLNHLNQPQKDALIQ 6432
QY 182 MAHPASA-----NAGDRLQHSPPHIPGSHHEIK----- 209
Db 6433 AINGATSRDQVAELKEAEALDEAMKQLEDQVODQDINSPPFINEDSDKQTYNDKIQ 6492
QY 210 --EPPVSGTSKAT-----TAHADRVETIAQEDDDSEFQOLHQORLARE 249
Db 6493 AAKEIINTSNPTLDKQKIADTLQNIKDANVNLHGQ--KLAQSKQDANNQNLHLLDLE 6551
QY 250 RENPPQPKLGVAIPISARFQPKLTAVAESVLEGTDTTQSPKPKQSMKLGSGAGVTPAV 309
Db 6552 QKN-----HFKEP-----LNNADTRDYNK----- 6571
QY 310 TLDGKQLQAPNPPALNTLTKGDTQHYLAHASSDGSQHLLLDNKGHLFDIKSTA 369
Db 6572 -----QLEIAQLNGDMSTLHKVINDKQIQIQLHSYINADNDKKQYDN-----AIK 6618
QY 370 TSYVLHNSHPGEIKGLAQACTGVSVDGKSGKISLGSQTSHNKTMLSPGFAHRSIL 429
Db 6619 EAEDLIHN-HPDTHKALQ-----DLINKIDQAHNEL----- 6650
QY 430 TGIWOHPAGAAR-----PQGESIR-----LHDDKIHLHPGLVWQS 466

Db 6651 -----NGESRFKQALDNALNIDSINLSNLNVPQRTVKDNINHVTTLESIAQEL----- 6698
QY 467 ADKDTHSQLSRQADGKIYALKD-----NFTQLNLSDNKSEKVLVDIKSY--- 511
Db 6699 -----QKAKELNDAMKARDSDIMNQEQIRKNSNYTNEDLAQONAYNHAVDKINNIIIGE 6751
QY 512 -----SVQORGQVAILTDTPGPHKMSIMPSLD 538
Db 6752 DNATMDPQIITQATQDINTAINGLNGDKLQDAKTDAKQOITNFTGLTEPQKQAL----- 6806
QY 539 ASPESHISLSLHFAHAGHGLHCK--SELEASQVAISHGRLVAD-----SEGLKF 587
Db 6807 -----ENTIINQOTSANVAKQLSHAKFLNGKMEELKVAVAKASLVQRNSYINEDVSEKAY 6863
QY 588 SAAIPKQGGCNE-LKMKAMPQHALDEHFGHDHQSIGFFHDDHGLNALVKNFRQOAHCP 646
Db 6864 EQAIAK-----GOEIIINSENNPTISSTDINRTIQEINDAEQNLHGD-----NKLRAQAEI- 6913
QY 647 LGNDHQFHPGWNLTDAIVDNQILGLHHTNPEPHEILDMHGLSLALQEGKLYFDQLTKG 706
Db 6914 AKNEIQNLGLNSAQITKLQIDIGRTTPKPAVTKLEAKAINQAMQOLKQSIADK----- 6969
QY 707 WTGAESDCKQLKKGDLGAAVLLKDGVEVKRLNIN-----QSTSSIKHGTENVFSLPHVN 760
Db 6970 -----DATLSSNYLNEDSEKKLAYDNVSAEQALINQNLNDPTWDISNIQAITQ 7018
QY 761 KP-EPGDALQGLNKKDKAKAMAVIGVKNYLAALTEKGD-----IRSFQIKPGTQQLERPA 813
Db 7019 KVIQAKDSLHGANKLAQNAQADSNLIINOSTNLNDKQKQALNDLNHNAQTKQQAIEIIAQA 7078
QY 814 QTLISREGISGELKDIIHVD-----HKQNLIALTHEGEVHQPREAQNGAESSSW----- 862
Db 7079 NKLNNE--MGTLLKTLVEEQSNVHQSKY--INEDPQVQNIYNDSTQKREILNGTTDDVLN 7135
QY 863 -HKLALPQSESCLKSLDMSHEHPIATFEDGSQH-----QLKAGGWHAAYAAPERG 911
Db 7136 NKKIADAIQNIHLTKYNDHGDQKLOKAAQDATNELNLTNLNNSQROSEHDEINSAERT 7195
QY 912 PLAVGTSGSQTIVNRLMGVKGKVIKPGSLTVKLSAQGTGAGTGAEGKRVSKFKSERIRAY 971
Db 7196 EVSNDLNHAKAL-NEAMQLENEVALENSVK-KLSDFIN-----EDEAAQNEYSNALQKA 7248
QY 972 --AFNPWTST--PRPTKNAAYATQHWQCGREGKPLIYEMQGALIKOLD--AAN----- 1018
Db 7249 KDIINGVPSSTLDKATTEDALLEQNARESLHGQKLOEAKQNAVAEIDNLQALNPGQVL 7308
QY 1019 -----VRHNAPOPDLQSKLETLDLGEHGAELNDMKFRDELEQSAIRSVTVLGQHQGV 1073
Db 7309 AEKTLVQASTKPBVQAL-----QKAKELNEAMKALTYINKK-----EQIK 7351
QY 1074 KSGEINSEFPKSPFKALVQSFN--VNRSGQ-----DLKSLQQAQVHAATPPSAESKL 1123
Db 7352 ADSRYVNAD-----SGLQANYSALNVGSIATTPPELNDKDVINRATQITKTAENNL 7405
QY 1124 --QSMLGHFVSAG--VDMSHQGEIPLGRQDPNDKTALTKSRLILDVTIGELHELADKA 1180
Db 7406 NGOSKLAESAQSDGNSQSTLEHQQ-----LTQSQK-----DKQHDNLNQA 7443
QY 1181 KLVSDHXPADQI-----KQLRQOQFDTLREKRYESNPVKHYTD-----MGFTNKALEA 1229
Db 7444 QT-----KQOVDIIVNNSKQLDNSMNQLOQIIVNNDTVKQNSDFINEDSSQDQDAYNHAQ 7499
QY 1230 NYDAVKAFINAFKKEH-----HGVNLITR----- 1253
Db 7500 AKDLITTAHTPTIMDKNQIDQAIENIKQALNDLHGSNKLESEKKEASEQQLNLSLNTNGQD 7559
QY 1254 TVLESQCSAEL-----AKKIKNTLLSL-----DSGESMGFSRSGYGGGVTFVPT 1298
Db 7560 TILNHIIPATRSQVGEKIASAKQNLNTWKALRDSIADNNELLOSKKYFNEDESEQ----- 7614
QY 1299 LSKKVPVPIPCAGITLDRAYNLSFRTSGGLNVSFGRDGGVSGNIMVATGHDVWPMYTG 1358
Db 7615 -----QNAYNQAVNKAKNIN-----DQPTPVMANDEIQSVLNE 7648

```

QY 1359 KTSAGN-----ASDWLSAKHISPDLRIGAAVSGTLQGTIONSLK-----1399
Db 7649 VKQTKDNLHGDKLANDKTDQAATNALYNLAQAGNLETKVQNSNRPVQKVQVLAN 7708
QY 1400 -----FKLTEDLPFGIHLGTHGLTPABELLOKHGHOMKQSKLTFSDVTS--ANLD--1450
Db 7709 QNDAMKLLDDALG-----NDAIKQTSNY--INEDTQQVNFDEY 7747
QY 1451 LRAGINLNEDEGSKPVGVTARVSA-----GLSASANLAAGSRERSTTSGQFGSTTS 1500
Db 7748 TDRGNIIVAEQTPNPNMSPNINTIADKITEAKNDLHGVOKLQAQOOSINTINQMTGLNQ 7807
QY 1501 ASN-----NRPTFLNGV--GAGANLTAALGYAHSTHEGKPVGIF 1538
Db 7808 AQKEQLNOEIQOTRSEHVQVINKAALNDMSMTLRQSIITDEHEVKQTSYINIEVGNQ 7867
QY 1539 PAFSTNTVSAALANDNRFSQSI--SLEKRAEPTVSN--DISLTSTLGHKHFKDSATTML 1595
Db 7868 TAYNNA--VDRVKQIINQTSNPTMNFLEVERA--TSNVKISKDALHGERELNDKNKSTF 7923
QY 1596 AA--LKELDADKAPAEQLHLOQHF-----SAKQVVG-----1624
Db 7924 AVNHLNLAQAKKEALTHEIEQATVSVQNNIYNKAKALNDMMKKLDLVAQODNVROSN 7983
QY 1625 -----DERYEAVERNKLKLVIROQAAD--SHSM 1649
Db 7984 NVINEDSTPQNMNDTINHAQSIIDQVANPTMSHDEIENAINIKHAI--NALDGEHL 8040
QY 1650 ELGSASHSTTYNNLSRIN--NDGIVELLHKKHFDALPASSAKRLGEMNNDPALKDIIK 1706
Db 8041 QOAKENALLNSLDLNAPODAJNLVNE-----AQTRKVAEQLQSAQALNDAMK 8093
QY 1707 QLQSTPFSSASVSMELK-----DGLREOTEKAILQGVREBVGVLFO-----DRNNLRV 1756
Db 8094 HLRNSIQNSSVRQESKYINASDAKKEQYNHAV-----REVENIINQHPFLDKEIILQ 8147
QY 1757 KSVSVQSVKSEGNTPALLIGTSNSAAMS-----MERNIGTFINPKYQDQNTPRFRFLE 1812
Db 8148 LTDGVNQA--NNDLNGVELLDADKQNAHQSIPTLMHLNQAQONALNEKINNATRETEVA 8204
QY 1813 GGIAQANQVASALATLKKEGLE 1835
Db 8205 AIIQOAK-----LLDHAVENLE 8221

RESULT 12
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12610

Query Match      2.7%; Score 257; DB 9; Length 5795;
Best Local Similarity 18.4%; Pred No. 8.3e-08;
Matches 377; Conservative 326; Mismatches 878; Indels 468; Gaps 91;

QY 16 TAAHNPVGHVALQQ---GSSSSSPQNAASAALAEKGRKMPRIHQBPSTAADGISAHQ 72
Db 2647 TAAQRVIDNGDATAQOISIDENTTAQRNDLTNQISQATNLAAVESVKQSANSLDG-AMGNL 2705
QY 73 QKKSFLRCGLGTTKXFSAPQCGPQGTTHSKGATLRDLLARDGGTQHEAAAPDARLTR 132
Db 2706 QTAINDKSGTLLASQNF--LDADQEKENAYNOAVSNAETILNKQTGPNTAKTAVEQALNNVN 2764
QY 133 S-----GVKRRNMDMAGRPWKGS-----GEDKV-----PTQQRKHOL 168
Db 2765 SAKHALNGTQNLNNAKQAAITAINGASDLNOKQDKALKAQANGAQRVNAQDVORNA TEL 2824
QY 169 NN-FQOMRQTMLSK---MAHPASANAGDRLOHS--PPHIFGSHHEIKKEP--VGSTSKATTA 222
Db 2825 NTAMGTLKHAIDAKTNTLASSKYVNADSTKQNAVTKVTNAEHIISGTPVTVTSEVTA 2884
QY 223 HADRVETIAQEDDDSEFQOLHQORLARENPPOPKLGVATPISARFQPKLTAVAESVLE 282
Db 2885 AANQVNSAKQELNGD-----ERLRAKQNAANTAI DALTQLTNPQAKLKEQVQGA--NBLE 2938
QY 283 GTDTTQSPKLPQSM-----LKGSGAGVTPLAVTLDKGLQAPDNPALANTLLKQTLG- 335
Db 2939 DIQIVQT--NQALNNAMKGLRDSIANET--TVKASQNYTASPNQSTYSAVSNAGKI 2994
QY 336 -----KDTQHYLAHASSDGSQHLLLNKNGHLFDIKSTA-----TSYSLVHNSH 379
Db 2995 INQTNPTMDTSAITQATTQVNNAKNG-----LNGAENLRNAQNTAKQNLNLTSLHTNNQ 3049
QY 380 PGEIKGLAQAGTGSVSDGSKISLQSGTGQSHNKTMLSQDGEAHRSLLTGIWOHPAGA 439
Db 3050 KSAISSQIDRAGHVSEVTAAKNAATEL--NTQMGN---LEQAIHDQNTVKQGVNFTDADK 3104
QY 440 ARPQGESIRLHDDKIHILHPELGYVWQSAKDQTHSOLSQAQDKLYALKDNRTLNLSDNK 499
Db 3105 AKRDAYTNVSRSET--ILNKTTQGA-NTSKQDVEAAIQNTVSATK-NALNGD---QNVWNAK 3158
QY 500 SSEK-LVDKIKSYVDQGVAILTDTFGRHKMSIMPDLSDASPESHISLSLHFADAHQGL 558
Db 3159 NTAKHALLNLTSLINNAQKRDLTTKID-----QATTVAGVEAVSNTGTQLTAMANLQNGI 3213
QY 559 LHGKSELEAQSVALSHGRLVADSEK-LPSAAIP-----KQDGNELKMKAMPQHALD 611
Db 3214 NDKANTLASENY---HD-----ADSDKKTAYTQAVTNAENILNKNSGSLD-KAAAVENALS 3265
QY 612 EHFCHDHIQSGFFHDDHQLNA-----LVKNPFQO-----642
Db 3266 QVTWAKGALNGNHNLEQAKSNANTTINGLOHLTTAQDKLKKQVQQAQNVAGVDTVKSSA 3325
QY 643 -----HACPLGNHQ-----FHPGMNLTAL-----VIDNOLGLHHTNPEPEHILDM 684
Db 3326 NTLNGAMGTLRNSIQDNTATNNGQNYLDATESNKTNYNNAVDSANGVINATSNFN--MDA 3383
QY 685 GHLGSLALQSGKLHYFDQLFKGWTGAESDCKQLK-----GLDGAAYL-----LKDG-----731
Db 3384 NAINQIATQVTS-----TKNALDGGTHNLTOAQKATNAIDGATNLNKAQKALKAQVT 3436

```


; NAME/KEY: MISC FEATURE
 ; LOCATION: (328)..(328)
 ; OTHER INFORMATION: X=any amino acid
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (336)..(336)
 ; OTHER INFORMATION: X=any amino acid
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (350)..(350)
 ; OTHER INFORMATION: X=any amino acid
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; LOCATION: (362)..(362)
 ; OTHER INFORMATION: X=any amino acid
 ; US-10-282-122A-70176

Query Match 2.7%; Score 256.5; DB 12; Length 2398;

Best Local Similarity 19.0%; Pred. No. 2e-08;

Matches 414; Conservative 289; Mismatches 866; Indels 613; Gaps 98;

QY 7 GTEHKAHVTAHNPVGHGVALQGGSSS-----SSPQNA---SLAAEGK 49
 DB 468 GQONLQRAKTEATNALTHASDLNQAKNALTOQVNSAQNVQAVNDIKQTTQSLNTAMTGL 527
 QY 50 NRKMPRIHOPSTAADGISAAHQOKKFSLSRCLGCTKKFSRPAQOQGTTHSKGATLRD 109
 DB 528 KRGVAN--HNQVQSDNYNADTNKKNDYNNAYNHANDIINGNAQHPVITPDSVNNALSN 585
 QY 110 LLARD--DGETOHEAAAPDA---ARLTRSGVVERNMDDMAGRPVKGSGGDEKVPF- 161
 DB 586 VTSKEHALNGEAKUNAKQEAENTALGHLNUNNAQONL-----QSQINGAHQIDAVNTI 640
 QY 162 QOKRHQLNN-FQGMQRTMLSK-----MAHPASANAGRLQHSPPHI--PGSHHEI 208
 DB 641 KQNAITNLNSAMGNLRFQAVADKQVKETEDYADADTAKQNAVNSVSSAETIINQTTNPTM 700
 QY 209 KEEPVGTSKATTAHADRV---EIAQEDDDSEFOQLHQRLARENPP---QPPKLGV 261
 DB 701 SVDDVNRATSAVTSKNALNGVEKLAQSKTDA-----ARAIDALPHLNAQAQKADV 750
 QY 262 ATPISARFPQKLTAVESVLGTDIT---EIAQEDDDSEFOQLHQRLARENPP---QPPKLGV 308
 DB 751 KSKINA--ASNIAGVNTVQOGTDLNTAMGNLQGAINDQETTLNSQNY-----QDAPTSK 803
 QY 309 VTLDPKGLQLAPDNPALNTLLKQTLGKDTQHYLAHASS-----DGSQHLILLDNKGHLF 363
 DB 804 KTAVTNAVQAADK--ILNKSNGQNKTKDQVTEAMQVNSAKNLDGTR--LLD-----852
 QY 364 DIKSTATSYSVIHN-SHPGEIKGLAQAGTGSVVDGSKGKISLGSGTQSHNKTMLSPQ 422
 DB 853 --QAKQTAQOQLNMWTH-----LTTAQKTNLTNQINSQTTVAGVQTVQSNAMTLDQAM 903
 QY 423 EAHRSLLTGIWHPAGARPQGESIRLHDDKHIL-----HPELG---VW 464
 DB 904 NTLRQSTAN-----XDATKASEDYDANNKQTYNNNAVAAAEIINANSNPNPSTIT 958
 QY 465 QSADKQTHSLSRQADGKLYALKDN-----RTLQNLSDNKSSEKLVDKIKSYS-----VDQ 515
 DB 959 QKAEQVNSKKTALNGDENLAAAKQNAKTYLNTLTSITDAQKN-NLISQITSATRVSGVDT 1017
 QY 516 RGQVAIITDTPGRHMSIMPOLD--ASPESHISLHPADAHQGLLHGKSELEAQSVASIS 573
 DB 1018 VKQNA-----QHLDAQMASLQNGINNESQVKSSEKYRDA-----DTNKQEQYDVAIT 1064
 QY 574 HGRVLVADSEG-----KLFSAAI-----PKQGDGNEIKMKAMPOHAIHDEHPGHQISGF 623
 DB 1065 AAKAILNKSTGPNTAQNAVEAALQRVNNAKDALNGDAKLIA-AQNAAKQHLG-----TL 1117
 QY 624 FHDDHGQLNALVKNFRQQHACPLGNDHQHPGPNWLTDLVIDNQGLGHLHTNPPEHILD 683
 DB 1118 THITTAQRNDLT-NQISQ--ATNLAVGESVKQANSIDGAMGNLQTAINDKS-----1166

QY 684 MGHLSLALQEGKLYHFDQITKGWTAESDCKQKKGLDGAAYLLKDGVEKRLINOSTS 743
 DB 1167 ----GTLASQ----NFLD-----ADEQKNAYNQAQVAAETILN-----KQCGPNTAKT 1207
 QY 744 SIKHGTENVFSILPHVRNKPEPGDAIQGLNKDDKAQAMAVIG---VNKYIALALTEKGBIRSF 800
 DB 1208 AVEQALNNVNNAKHALN-----GTQNLNNAQAAITAINGASDLNKKQKDALKAQANGA 1261
 QY 801 QIKPGTQOLERPAQTLRSIGISGELKD IHVDH-----KQNLV-----837
 DB 1262 QRVSNAQDVQHNA TELNT--AMGTLKHAIADKNTINTLASSKYVYNADSTKONAYTTKVTNAE 1319
 QY 838 -----ALTHEGEVF-----HOPREAWQNGAES-SSWHKIALPQSSS 872
 DB 1320 HIIISGTPVTVTPSEVTTAAANQVNSAKQELNGDEHREAKQNAITAI DALTOINTPQKAK 1379
 QY 873 KUKSIDMSHEHKPIATFFEDGSOHLKA-GGHAYAAAPERGLAVGTS-----GSQTV 923
 DB 1380 LKEQVGOANRLLEDVQVTNGQALNNAMKGLRDSIANE---TTVKTSQNYTSDASPNQST 1436
 QY 924 FNRLMQGVKGVIPGSGLTIVKLSAQGTGNTAGRKVSKFSERIRAYAFNPTMTSPRI 983
 DB 1437 YNSAVNAKGIINQTNNTMTDSAITQATTQVNNAKNGLNGAENLR-----NAQNTAKQNL 1492
 QY 984 KNAAYATQHGMOGREGKPLYEMOGALIKOLD--AH-----NVRHNAPOPDLQSKLETLDL 1037
 DB 1493 NTLSHLTNN-----QKSAISSQIDRAGHVSSEVATKNA-ATELNTQGNLEQ 1538
 QY 1038 GEHGAELLNDMKRF-----RDELEQSATRSVTVLGGHQVLKSGNINSEFK--PSPG 1088
 DB 1539 AIHDONTVKQSKFTDADKAKRDAYTNAVSRAEAILNKTQGAITSKQDVEAAIQNVSSAK 1598
 QY 1089 KALVQSENV-----NRSQDLKSLSQOAVHATPPSAESKLQMLGHF 1130
 DB 1599 NALNGDQNVTNNAKNAKNALNLT SINNAQKRDLTTKIDQATTVAGVEAVSNTSTQLNTA 1658
 QY 1131 VSAGVDMSHQGEIPLGRQRPNDKT-----ALTKSRLTLD---1166
 DB 1659 MA-----NLQNGINDKNTLASENVHDADSKKTYATQAVTNAENLTNKS 1704
 QY 1167 ---TVTIGELHELAD-KAKLVSDH-----KPDV-----DOIQLRQQ 1199
 DB 1705 GSNLDKTAVALNSQVANAQKALNGNHNLEQAQSNANTINGLQHLTTAQDKLQKQVQ 1764
 QY 1200 -----FDTLREKYESNPVGHYDMGTGNKALE-----ANY-DAVKA-----FINAF 1241
 DB 1765 AQNVAGVDTTVKSSANTLNGA-----MGLTRNSIQDNTATKNGQNYLDATERNKNTYNNAV 1819
 QY 1242 KKEHHGVNLTTRTVLESQGSAAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSK 1301
 DB 1820 DSANGVINATSNPNWDANAIQIATQVTSITKNALDGTNHL-----TOAK 1863
 QY 1302 KVPVPIPGAGITLDRAYN-----LSFSRTSGGLNVSGF-----1335
 DB 1864 QATNAIDGA-TNLNKAQKADKALKAQVTSQKRVANVTSIQOTANELNTAMQGLQHGDIDEN 1922
 QY 1336 -----RGQVSGNI---MYATGHVMPYMTGKTSAGNASDWL-----SAKHKTSPP 1379
 DB 1923 ATKQTKYRDAEOKKTYAYDOAVAAAKAILNKQTSNSDKAAVDRALQVVTSTTDALNGD 1982
 QY 1380 LRIGAAVSGTLQ-GTLQNSLKFKLITDELPGFTHGLTHGLTLPAAELLQKGIHQHMQGS 1437
 DB 1983 AKLAEAKAAKQNLGTLNHTNAQRTD-----LEGQINQAT 2018
 QY 1438 KL--TFSVDTSAN-LD-----LRAGIN-----LNEDGSKPNGVTVARVSAGLSAS 1478
 DB 2019 TVDGVNTVKTNANTLDGAMNSLQGSINDKATLRNQNYLDADESKRNAYTQAVTA-----2073
 QY 1479 ANLAAGSRRESITSGQPGSTTSANRPFTLNGV-GAGANILTAALGVVAHSSSTHEGKPVGI 1537
 DB 2074 ---AEGILNKQTG---GNTSKADVNA--LNAVTRAKAALNGADNLRNATNTIDG 2124
 QY 1538 FPAPTSTNVSAALALDNRSTQSISLELKRAEPVTS-NDISELSTLGLKHKFKDSATTQWLA 1596

Db 2125 LPNLTQ-----LQKN-----LKHQEQAVGVGVKDKGNTL-----NTMAGALRT 2168
 QY 1597 ALKELDDAKPAEQHLHQHFSKQVGVDERVEAVNLKLVIRQQAADSHSMELGSASH 1656
 Db 2169 SIQNDNTTKTS-----QNYLDASDNKNYNTAVNANGVI-----NATNPNNDANAING 2219
 QY 1657 STTYNLSRINDGIVELLHKKHFDAAIPASSAKRLGEMMNDPALKDIIKQLQ----- 1709
 Db 2220 MANQVNTTKAALNGAQNLAQAKTINATNTINNAHDLNKKQKD--ALKTVNNAQVSDANN 2277
 QY 1710 ----STFFSASVSMELDKGLRQGT-----EKALDGKVGREVGVLFODRNN 1753
 Db 2278 VOHTATELNSAMTALKAAIADKERTKASGNVYVADQEKQAYDSKVTVNAENIISGTPNAT 2337
 QY 1754 LRVKSVSVQSQSVKSGFNTPALLLGTSNSAASMERNTGTFNKFGYQDQNTPRRTLEG 1813
 Db 2338 LTVDVN-----SAASQVNAAKTALNGDNNLRVAKHEANNTID----- 2375
 QY 1814 GIAQAN-----POVASALT 1827
 Db 2376 GLAQLNNAQAKLKEQVQSATT 2397

RESULT 14

US-10-282-122A-74463
 ; Sequence 74463, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 74463
 ; LENGTH: 2045
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pyogenes
 US-10-282-122A-74463

Query Match 2.7%; Score 253; DB 12; Length 2045;
 Best Local Similarity 18.9%; Pred. No. 2.8e-08;
 Matches 418; Conservative 308; Mismatches 785; Indels 706; Gaps 100;
 QY 22 VHGVALQOQSSSS-----PQNAASALAAEGKNGKMPRIHQSTAAADGISAHH 71
 Db 17 VGLGVAVPTGFGSQSNGVMVKAAEVPAATLRSQASDSERVDSSLLQKENLSDV----- 70
 QY 72 QOKKFSRLGCLGCTKKFSRPAQOGPTTHSKGATLRLDILLARDGGTQHEAAAPDAARLT 131
 Db 71 ----SKLENLNGWEAENDTA--GNLGKFKDPDSSGVQNILTSSGKNISVAVAPK----- 119
 QY 132 RSGGVKRNMDMAGRPVMKVGSGGDKVPTQKXKHQNLNP-----GQVRQTMLSKVAH 184
 Db 120 ----GSGKNIKVTKR---SNFGGYVVGGLRTQTPVLKL-- 152
 QY 185 PASANAGDRLQHSPPHPPGSHHEIKE--EPVGTSK-----ATTAAHADVET 229
 Db 153 ----NDVYRYSFTTKKLSGNSSEFKTRVVPESNNKLGKELVIRVNKNVSTKHDMLPDI 208
 QY 230 AQEDDDSEFOQLHQORLARERENPPQPKLGATPISARFPKLT----- 274
 Db 209 SDGTHVTVDFTGLDK-----KLSVA---FRSPQTSNVVYEFNSINIKNI 250
 QY 275 ----AVAESVLEGTDTTQSPKPKQSMKLGSGAGVTPLAVTLDKKG-----LOLAPDN 322
 Db 251 SPASVPAIPSKVLEGT-----SVLSGTAIS---SGDTLEKRSFGDILRVYKDS 297
 QY 323 PPALNTLLKQT-----LGKDTQHYLAHASSDQSHLLDNKGHFLDFTKS--- 367
 Db 298 KIARTVTKGNKNDVKLSKPLIAGEKLDPEILHPRSONVSKISKQVEAFPPASVKEK 357
 QY 368 ----TATSYSVLHNSHPGE-----IKGKLA--QAGTGSVSVDSKSG 402
 Db 358 VIAKLFPVYEATSEKITNDAWLDENAKDLQKLEQYISGKVAISEAGTKQEAIDAAYN 417
 QY 403 KI-----SLGSGTQSHNKTMLSQPG-----EA 424
 Db 418 KYSSQTDPSQYKQGNKENEQEGRODLIOTRDLTLKAIQEDKWLTEQKTIQKEEA 477
 QY 425 HRSLLTGIV-----OHPAG-----AARPP 443
 Db 478 LKAFETGIESVNTVSLEQLKQRLIVYKASEKSEKYEPIESIPNQHIPGKEVKAQAK 537
 QY 444 GESIRLHDDKIHLHPELVGMQSAKDTSHLSQADGK--LYALKONRTLONL----- 495
 Db 538 EELKLDHTTLEKINQD--KWLTPDQAEQLQAEVTFKKQEAIAKSAQTLTQLETDLAD 595
 QY 496 --SDNK--SSEKLVDKIYSVD---QRQVAILTDTGPRHKMSIMPSLDASPEHSLSL 549
 Db 596 YVSENEGKNSIPDKYKSGNKKDLVNKAEVK--LKEAHEATKQAIKEDPWLSPQK----- 649
 QY 550 HFADAHQGLLHGKSELEAQSVASHG--RLVVADS-----EGKLFSAAPKQ-- 594
 Db 650 ----KAQKEKAKARLDEGLKALKAAADSLEILKVTEAFVYDKKNPDSIPNQH 697
 QY 595 --GDGNELKMKAMPQHALDFHGHQISGFHDDHGLQNALVKNVNFQQHACPLGNHQ 652
 Db 698 KAGTADQARKQAL--DSLDK-----EVQKELESIDNDN-----TLTTDEK 735
 QY 653 FHPGWNLTDALVINQGLHHTNPEPHEILDGMHGLSLALQEGKLYHFDQLTGKWTGAS 712
 Db 736 AAARKKVNDAVDVAKOTAMEANSYEDLTTIKDEFSLNLFHQKQ--TPLKDDQSDAIAELEK 794
 QY 713 DCKOLKGLDGAAYLLKDKGEVRLNINOSTSSIKHGTENVSFLPHVR-----NK 761
 Db 795 KQIEIKAEIGDKTLPRDEKEKQ--IADSKELASDIQKVKADAKNADAIIKAFEBEGKNI 852
 QY 762 PE---PGDALQGLNKDKKAQAMA-----VIGNVKYALTE----- 793
 Db 853 POAHIPGD---LNK--DKEKLLAELKQKADDTKAIKADDTKTEDEKKEQKVKTKAELEK 907

Db 1329 DOVERLAQKQTAGNSLHLDLTTPAQOQALENOINNATTRDKVA-----EIIA 1377
 QY 276 VAESYLEGDTTQSPKPKQSM.KGSGAGVTPPLAVTLD--KGKLQAPD--NPPALNTLLK 331
 Db 1378 QAQALNEAMKALKEGSIKDPQTEASSKFINEQAQKDAYQAVQAHKDLINKTTPTLVK 1437
 QY 332 QTLGKDTQHYLAHSHASSDQSHL-----LLDNKGHLFDIKSTATSYSLVHNSHPC 382
 Db 1438 SVIDQATQAVNDKNNHLDQKLAQKQATETLNLNSLNTPQOALENOINNATRGE 1497
 QY 383 IKGKLAOAGCTSVSDGKSGIKSLGSGTQSHNKTMLSPQCEAHRSLLTGIGIWHPPAGAAR 442
 Db 1498 VAQKLEA-----QALNQAM-----EALRNSIQOQOQTEAG----- 1528
 QY 443 QESIRLHDKTHILHPELGVWQSA---DKDTHSQLSROADGKLYALKDNRTLNLNSDN- 498
 Db 1529 ---SKFINEDK-----POKDAYQAAVQAHKDLINQTSNPTLKAQVEQLTQGVNQAKDL 1580
 QY 499 KSEKLVADKIKSVSYDQRGQVAILTDPGRHKWMSIMPSLDASPESHISLHPADA-HQG 557
 Db 1581 HGDQKLAAD-KQHAVTDLNQLNSL--NNPQOALESQIN--NAATREVAQKLABAQALDQA 1637
 QY 558 LLHGKSELEAQAQSVAGHRLVADSEK-LFSAAIPOQGD-----GNELMKAMPQHAL- 610
 Db 1638 MQALRNSIQOQOQTESSKFINEKPKQDAYQAAVQAHKDLINQTSNPTLKDQSQVEQLTQ 1697
 QY 611 -----DHFHGDHQ-----ISGFPHDGHQNLAL-----VKNNFRQOHACP 646
 Db 1698 AVTTAKDNLHGDKLARDQOQAVTTNNALPNLHAQOQALTDAINAAPTREVAQHVCQA 1757
 QY 647 LGNDHQPHCPGNLTDAIVDN-OLGLHHTNPEPEHILDMGHGLSALQEGKHYFQDLTK 705
 Db 1758 TELDHAMETLKNVQVNTDKAQNYTEASTDKKEAVDQ-----ALQAA-----ESITD 1806
 QY 706 GWTGABSDCKQLKGLDGAAYLLKDGVEKRLINQSTSSIK-HGTENVFSLPHVRNKPPEP 764
 Db 1807 PTNGSNAN---KDAVEQALTKQE-KVNELNGNERVAEAKAQAKOTIDQLAHL-NADQI 1860
 QY 765 GDALQGLNKDDKAQAMAVIGNKYALATEKGDITSFQIKGTQQLERPAOTLSREGISGE 824
 Db 1861 ATAKQNTIDQATKLQPIAEL-VDAQTLNQSDM---QLQOAVNEHANVEQTVDTYQADSD 1915
 QY 825 LKDIHVDBKQNLVALTHEGEVFHPREAWQNGAESSSWHKLALPQSEBSKLKSLDMSHEHK 884
 Db 1916 ---KQAY-----KQATAEAENVL-KQNSNKKQ-- 1939
 QY 885 PIATFEDGSOHLKAGQWHAYAPERGLAVGTSGSQT-----FNRLMOGVKKG----- 934
 Db 1940 ---VDQALQNLNAK--QALNGDERVALA-KTNGKHDIDQLNALNNAQDQGFGRIDQS 1992
 QY 935 -----VTPGSLTQKLSAQTCGTMGTAEGR-----KYSSKFSERI----- 968
 Db 1993 HDLNCIQOIVDEAKALNAMDQUSQEIISGNEGRTKGSNTYVYNADTVQKVYDEAVDKAKQ 2052
 QY 969 ---RAYAFNPTMTSPRTKNAAYATOHGWQREGKPLIYEMOGGALIKQLDAHNVHRNAPQ 1025
 Db 2053 ALDKSTGQNLTAQVILKINDAVTAAKKALNGEERLN---NRKSEALQRLDQLTHLNN-- 2107
 QY 1026 PDLQSKLETDLBGHGAELINDMKRFRDELEBQSATRSVTVLG-----OHQGVLSKNG 1077
 Db 2108 ---QRQLAIQOI---NNAETLNKASRAINR---ATKLDNAMGAVQVQIDEQHLGVISSTN 2158
 QY 1078 EINSB--FKPSPGKALVQSFVNRSGODLSLOQAVHATPPSAESKLOSLMG--HFVSA 1133
 Db 2159 YINADDNLKANYDNAIA-----NAAHELDKVVQGNAI---AKABAEQLKQNIIDQALNG 2210
 QY 1134 GVDMSHOKGEI-----PLGRQRPNDKTLTKSRILDLTVTIG-----ELHELADKAK 1181
 Db 2211 DONLANAKDKANAFVNSLNGNLQOQDLAHNAINNADTVSDVTDIVNNQIDINDAMETLK 2270
 QY 1182 -LVSDHKPDADQIKQLRQOQFDTLR-----EKRYESNPVKHYTDMGFTHNKALEANYDAV- 1234

2271 HLVNDNEIPNAQETVNYQNADDAKTNFDDAKELANTLNSDNTVNDINGAIQAVKDAIQ 2330
 1235 -----KAFINAFK-----KEHGVNLTTRTVLESQSA--ELAKKLKNTLLS 1274
 2331 NUNGQRLOEAKDKAIQNVKVLAKLKEIEASNATDQDKLIAKNAEELANSIINNINK 2390
 1275 LDGSGMSFSRSGGVSFVPTLSKKVPVPIPGAGITLDRAYNLSFRTSGGJNVSF 1334
 2391 ATSNQDVSVQTAGN-----QAIEQVHANEIIPRAKIDA---NKQVDKQVQALIDEI 2438
 1335 GRDGGVSG-----NIMVATGH-DWPMYMTGKTSAGNASDWLS----- 1371
 2439 GRNPNTDKKQALKDRINOIILQOQHNDINNALT--KEAIEQAKERLAQALQIDKLVKA 2496
 1372 ---AKEHI-----SPDL-----RIGAAVSGTLOGTTLONSILKFLKTED 1405
 2497 KEDAKNKIKALANAKEDQINSNPDLTPEQAKAKALKEIDEAEKRALQ-NVENA---QTID 2551
 1406 EL-PGPIHGL-----TH-----GTLTPAELLQKG--TEHOMKQSGKLTFSVD 1444
 2552 QLNRLGNLGLDDIRNTHVWVEDDQPAVNEISBATPEQLLVNGELIVHR-----DD 2601
 1445 TSANLDRAGINL-----NEDGSKPNGVTARVSAGLSA--SANLAAGSR-----E 1487
 2602 ITEQDVLAHINLIDOLTAEVIDTFS--TATISDSLTAKEVETLLDGSKVIVNVVPKWE 2659
 1488 RSTTSQFGSTTSASNNRPTFLNGVGAGANLTAALGVAAHSSHTHEGKPVGIFPAFTSNVS 1547
 2660 KELSVMVQQAIESIENAVQOKINEINNSVTTL-----EQK-----E 2696
 1548 AALALDNRTSOSISLELKRAEPTVS--NDISELTSTLGHKFKDSATTMKLA---ALKELDD 1603
 2697 AATAEVNKLKQQAIDHINNAPDVHVSVEIQOQEAHIEQFNPEOFTIEQAKSNAIKSIED 2756
 1604 AKPAEQHLHLOQHFSAKDVVDBRYEAVRNLKLVIRQQAADSHSMELGSAHSSTTYNNL 1663
 2757 AIO-----HMDEIKARTDLTDKEQERAKLNLQ--KEQA-----IQLIQAQSI----- 2800
 1664 SRINNDGIVELLHKHFDALPASS-----AKRLGEMMNNDPALKDIIKQLQSTPFPSSAS 1717
 2801 ---DEITEQL-EQFKAQMAANPTAKELAKRQEAISK---IKDFSNEKMNISRSEI 2851
 1718 VSMELKDLREOTEKAILDGVGVEEVGLFODRNN-----LRKVSVS- 1761
 2852 GTADEKQAAMNQINEIVLE-----TIRDINNAHTLOQVEAALNGIARISAVQIV 2901
 1762 ---SQSVKSGEFGNTPAL--LGTNSAAMSMERNIGTINFKYQDQON 1804
 2902 TSDRAKQSSSTGNSHSLTIGYGTANHPF-----NSSTIGHKKKLLDED 2945

Search completed: July 7, 2004, 15:30:44
 Job time : 105.321 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:14:07 ; Search time 31.6095 seconds

Title: US-09-596-784-2
Perfect score: 9448
Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTDLKKEGLEMK 1838
5593.254 Million cell updates/sec
(without alignments)

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 T18448	pathogenicity factor
2	1824	19.3	1795	2 T30332	avirulence protein
3	275.5	2.9	6713	2 B89921	hypothetical prote
4	275	2.9	2514	2 F81045	hemagglutinin/hemo
5	258.5	2.7	2248	2 A35938	profilaggrin - hum
6	255.5	2.7	3455	2 B82519	hemagglutinin-like
7	247	2.6	3442	2 B82589	hemagglutinin-like
8	245	2.6	2703	2 H81193	hemagglutinin/hemo
9	244.5	2.6	3259	1 A56539	giantin - human
10	243.5	2.6	3225	2 I52300	giantin - human
11	241	2.6	2535	2 AC0304	probable hemolysin
12	240.5	2.5	5327	2 T13564	microtubule-associ
13	237.5	2.5	4152	2 T31102	filamentous hemagg
14	236	2.5	2271	2 F90073	hypothetical prote
15	232	2.5	1577	2 A35140	hemolysin A precu
16	229.5	2.4	2273	2 T09083	hemagglutinin/hemo
17	221.5	2.3	1635	2 A10452	hemolysin (importe
18	221	2.3	1608	2 A28182	hemolysin A - Serr
19	220.5	2.3	2253	2 T30336	nuclear/mitotic ap
20	220.5	2.3	3295	2 AE0074	probable adhesin Y
21	220	2.3	3890	2 C89921	hypothetical prote
22	219.5	2.3	2541	2 S11661	talin - mouse
23	218.5	2.3	4957	2 T03455	ALR protein - huma
24	218.5	2.3	5262	2 T03454	ALR protein - huma
25	217.5	2.3	1957	2 T38077	hypothetical coile
26	217	2.3	2712	2 T05113	hypothetical prote
27	214	2.3	3282	2 E82750	hemagglutinin-like
28	212.5	2.2	5627	2 C83339	hypothetical prote
29	210.5	2.2	1788	2 AH1447	probable tape-meas

30	210	2.2	2059	2 D82671	surface protein XF
31	209.5	2.2	1510	2 T31100	hypothetical prote
32	209.5	2.2	4919	2 T31105	hypothetical prote
33	207.5	2.2	3591	1 S21010	filamentous hemagg
34	205	2.2	1684	2 JW0057	gravin - human
35	204.5	2.2	4558	2 C82199	RTX toxin RtxA Vcl
36	204	2.2	2845	2 I49505	adenomatous polyo
37	203.5	2.2	1995	2 G81044	hemagglutinin/hemo
38	203.5	2.2	3187	2 JC5837	hypothetical prote
39	202.5	2.1	2218	2 B84683	hypothetical prote
40	202	2.1	1975	2 B81192	conserved hypochet
41	200	2.1	1467	2 A75564	hemagglutinin/hemo
42	199	2.1	2232	2 T34434	hypothetical prote
43	198.5	2.1	835	1 I57441	involucrin - orang
44	197.5	2.1	2015	2 B81989	hypothetical prote
45	196.5	2.1	1902	1 B44858	lactocepin (EC 3.4

ALIGNMENTS

RESULT 1

T18448
Pathogenicity factor DspA - Erwinia amylovora
C:Species: Erwinia amylovora
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T18448
R:Gaudriault, S.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z18936
A:Accession: T18448
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1838 <GAU>
A:Cross-references: EMBL:Y13831; PIDN:CAA74156.1
C:Genetics:
A:Note: dspa
C:Function:
A:Description: involved in pathogenicity

Query Match	100.0%;	Score	9448;	DB	2;	Length	1838;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	1838;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MELKSLGTEHKAHVHTAAHNPVGHGVALQOGSSSSPQNAASLAAGKNCMPRIHQP	60				
Db	1	MELKSLGTEHKAHVHTAAHNPVGHGVALQOGSSSSPQNAASLAAGKNCMPRIHQP	60				
Qy	61	STAADGISAAHOOKKSFSLRGCLGKTKFSRSAPQGPQGTTHSKGATRLDLLARDGGETQH	120				
Db	61	STAADGISAAHOOKKSFSLRGCLGKTKFSRSAPQGPQGTTHSKGATRLDLLARDGGETQH	120				
Qy	121	EAAAPDAARLTSGGVKRENMDMAGRPMVKGSGEDKVPTQOKRHQLNFFGQMQRTMLS	180				
Db	121	EAAAPDAARLTSGGVKRENMDMAGRPMVKGSGEDKVPTQOKRHQLNFFGQMQRTMLS	180				
Qy	181	KMAHPASANAGRLQHSPPHIFGSHHEIKKEEPVGSSTKATTAHADRAVEIAQEDDDSEFQQ	240				
Db	181	KMAHPASANAGRLQHSPPHIFGSHHEIKKEEPVGSSTKATTAHADRAVEIAQEDDDSEFQQ	240				
Qy	241	LHQORLARENNPPQPKLGAVATPISARFQPKLTAVAESVLEGTDTTQSPKPKQSMKGS	300				
Db	241	LHQORLARENNPPQPKLGAVATPISARFQPKLTAVAESVLEGTDTTQSPKPKQSMKGS	300				
Qy	301	GAGVTPAVLTLDKGLQAPNDPPALNTLLKQTLGKDTQHYLAHASSGSGSHLLLDNKG	360				
Db	301	GAGVTPAVLTLDKGLQAPNDPPALNTLLKQTLGKDTQHYLAHASSGSGSHLLLDNKG	360				
Qy	361	HLFDIKSTATSYSLVHNHSGEIKGLAQAGTGSVDGSGKISLGSGTQSNKNTLSQ	420				
Db	361	HLFDIKSTATSYSLVHNHSGEIKGLAQAGTGSVDGSGKISLGSGTQSNKNTLSQ	420				
Qy	421	PGENHRSLLTGIWQHPAGARPQGESIRLHDDKTHILHPGLGVQWSADKOTHSQLSRQAD	480				

Db 421 PGEAHSLLTGTWHPAGARPGESIRLHDDKHILHPELGWQVADKDTTHSQLROAD 480
QY 481 GKLYALKONRTTQNTSDNKSSEKLVKISYSVDQRGQVAILTDTTPGRHKMSIMPSLSDAS 540
Db 481 GKLYALKONRTTQNTSDNKSSEKLVKISYSVDQRGQVAILTDTTPGRHKMSIMPSLSDAS 540
QY 541 PSHHSLSLHFAHADAHQCLLHGKSELBAQVAISHGRLVVADSEGLFSAAIKQGGDNEL 600
Db 541 PSHHSLSLHFAHADAHQCLLHGKSELBAQVAISHGRLVVADSEGLFSAAIKQGGDNEL 600
QY 601 KMKAMPQHALDRHFHGHDIQSGFFHDDHGLNALVKNPNRQHQHACPLGNDHOFHPCWNLT 660
Db 601 KMKAMPQHALDRHFHGHDIQSGFFHDDHGLNALVKNPNRQHQHACPLGNDHOFHPCWNLT 660
QY 661 DALVIDNOLGHHHTNPPEHEILLDMHGLSLALQEGKLYHFDQLTGKWTGAESDCKQLKKG 720
Db 661 DALVIDNOLGHHHTNPPEHEILLDMHGLSLALQEGKLYHFDQLTGKWTGAESDCKQLKKG 720
QY 721 LDGAAYLLKDGVEKELNINQSTSSIKHGTENVFSLPHVRNKPPEPDGALQGLNKDKAQAM 780
Db 721 LDGAAYLLKDGVEKELNINQSTSSIKHGTENVFSLPHVRNKPPEPDGALQGLNKDKAQAM 780
QY 781 AVIGWNYKLYALTEKGDIRSFOIKPQTQOLRPAQTLRREGISGELKDIHVDHKQNYALT 840
Db 781 AVIGWNYKLYALTEKGDIRSFOIKPQTQOLRPAQTLRREGISGELKDIHVDHKQNYALT 840
QY 841 HEGEVFHPRAWONGAESSWHKIALPOSESLSKSLDMSHEHKPIATFEDGSHQKAG 900
Db 841 HEGEVFHPRAWONGAESSWHKIALPOSESLSKSLDMSHEHKPIATFEDGSHQKAG 900
QY 901 GWHAYAAPERGLAVGTSGSTVFNRLMOGVKGVIPGSGLTIVKLSAQTGMTGAERKV 960
Db 901 GWHAYAAPERGLAVGTSGSTVFNRLMOGVKGVIPGSGLTIVKLSAQTGMTGAERKV 960
QY 961 SSKFSERIRAYAFNPTMSTRPKIKNAAAYATQHGQGRGLKPLYEMOGALIKQIDAHNVR 1020
Db 961 SSKFSERIRAYAFNPTMSTRPKIKNAAAYATQHGQGRGLKPLYEMOGALIKQIDAHNVR 1020
QY 1021 HNAPOPLOQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLQHQGVLSKNGEIN 1080
Db 1021 HNAPOPLOQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLQHQGVLSKNGEIN 1080
QY 1081 SEKPSFGKALVGSNNRSGQDLKSLOAVHATPPSAESKLSQMLGHFVSAGVMSHQ 1140
Db 1081 SEKPSFGKALVGSNNRSGQDLKSLOAVHATPPSAESKLSQMLGHFVSAGVMSHQ 1140
QY 1141 KGEIPLGRQRPNDKTALTKSRLILDTVTIGELHELADKAKLYSDHKPDADQIKQROOF 1200
Db 1141 KGEIPLGRQRPNDKTALTKSRLILDTVTIGELHELADKAKLYSDHKPDADQIKQROOF 1200
QY 1201 DTLREKRYSPNPKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQ 1260
Db 1201 DTLREKRYSPNPKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQ 1260
QY 1261 SAEALAKKNTLLSLDGSMSFSRSYGGVSVFVPTLSKKVPVPVPCAGITLDRAIN 1320
Db 1261 SAEALAKKNTLLSLDGSMSFSRSYGGVSVFVPTLSKKVPVPVPCAGITLDRAIN 1320
QY 1321 LFSFRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPL 1380
Db 1321 LFSFRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPL 1380
QY 1381 RIGAAVSGTLOQTIONSLKFKLTEDLPFGFIHGLTHGTLTPAELLQKGIHOMQGSKIT 1440
Db 1381 RIGAAVSGTLOQTIONSLKFKLTEDLPFGFIHGLTHGTLTPAELLQKGIHOMQGSKIT 1440
QY 1441 FSVDTSANLDRAGINLNDGSKPNGTARVSGAGISASANLAAGSERSTTSQSGSTTS 1500
Db 1441 FSVDTSANLDRAGINLNDGSKPNGTARVSGAGISASANLAAGSERSTTSQSGSTTS 1500
QY 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Db 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
QY 1561 SLELKRAEPTVNDISELTSTLTKGPKDSATTMKAALAKELDDAKPAEQHILQQHFSAK 1620
Db 1561 SLELKRAEPTVNDISELTSTLTKGPKDSATTMKAALAKELDDAKPAEQHILQQHFSAK 1620
QY 1621 DVVGDERVEAVNLKGLVIRQQAASHSHSMELSGASHSTTYNNLSRINNDGIVELLKHFD 1680
Db 1621 DVVGDERVEAVNLKGLVIRQQAASHSHSMELSGASHSTTYNNLSRINNDGIVELLKHFD 1680
QY 1681 AALPASSAKRLGEMNNNDPALKDIIKQLOSTPEFSSASVSMELKDGIREOTEKAILDKGVG 1740
Db 1681 AALPASSAKRLGEMNNNDPALKDIIKQLOSTPEFSSASVSMELKDGIREOTEKAILDKGVG 1740
QY 1741 REEVGVLPQDRNNLRVKSVSQSVSKSEGFTNTPALLGTSNSAAMSERNITGTFNKYK 1800
Db 1741 REEVGVLPQDRNNLRVKSVSQSVSKSEGFTNTPALLGTSNSAAMSERNITGTFNKYK 1800
QY 1801 QDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838
Db 1801 QDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838

RESULT 2

T30332

C:Species: Pseudomonas syringae

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T30332

R:Bogdanove, A.J.; Kim, J.F.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.;

Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998

A:Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dspeF

A:Reference number: Z20825; MUID:98115919; PMID:9448330

A:Accession: T30332

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1795 <BOG>

A:Cross-references: EMBL:U97505; NID:g2978502; PID:g2978503; PIDN:AA06134.1

C:Genetics:

A:Gene: avrE

Query Match 19.3%; Score 1824; DB 2; Length 1795;

Best Local Similarity 29.0%; Pred. No. 8.9e-83;

Matches 555; Conservative 366; Mismatches 744; Indels 246; Gaps 72;

QY 55 PRH-----QSTADGISAHQKKSFSLRGCLGTFKFSRSAPQGGTTHSKA-- 105

Db 4 PSIHRTGSIQPTVTPDARAATDQER-----AEQPRQRSSHSLSSVG 47

QY 106 --TLRDL--LARDGGETOHEAAAPDAAR-----LTRSGGVKRRNMD-- 143

Db 48 KRAKSVGKLFQSKAPQQAATPTAKNVKTPPPASNVATPRNKARESGFSNSPQNT 107

QY 144 -----MAGRPYKGGSGEDKVPTQOKRHOLNFGQWOTMLSKMAHPASANAGDRLQHS 197

Db 108 RAPKWLIRNHPNQASSG-----AQTHEIPEAAPRKNLVRVFLPO-----DLERS 155

QY 198 PHIPGSHHEIKBBPVGTSKATTAHADREVEIAQEDDDSEFQQLHQRLARENPQP- 256

Db 156 PSYSDNDNPMTEEAANATROF-----RSPDLSHQSDGTRISMLATDPDQS 204

QY 257 ---PKLGAVT-PISARFQPKLTAVAESVLEGTDTTQSLPKQSMKKGSGAGVTP-LAVTL 311

Db 205 SSGSKIGSDGPIPR-----EPLWRSNGGRFELDKELVRNS-----EPOGSIOL 251

QY 312 D-KGKLQAPDNPPALNTLLKQTLGKDTQHYLAHSHASDGSQHLLDNKGHLFDIKSTAT 370

Db 252 DAKGKPDFSTFPGPLDLSILATPKQTYLAHQSKGVHGHQQLQANGHLHLAQQDS 311

QY 371 SVSVLHNSHPG-EIKGKLAQAGTSVSDGSKISLGSQGTOSHNTMLSQFGEAHRSL 429

Db 312 SLAVIRSSNEALLIEGKKPPA-----VKMEREDGNTHI--DTASGRKTQ-ELPGKAHIAHI 364

QY 430 TGIWQHAGARPGESIRLDDKHILHPGLGWQSD--KDT-HSLSRQADQKLYAL 486
 Db 365 TNV-----LLSDGERMRVHEORLYOFDPISTRWKIPEGELEDTAFNSLSTGGNGSYAK 418
 QY 487 KDNRTLQNLSDNKSSEKLVKIKSYSDQGOVAILLTDPFRHKMSIMPSLDASPEHSIS 546
 Db 419 SDDAVV-DLSPFMPHVEVEDLQSFVAPDNRAALLS---GKTTQAILLT-DMSF-----V 469
 QY 547 LSHFADAHQGLLHGKSELEBAQSAVISHGRIVADSEKLFSAAI PQOGDNELKMKAMP 606
 Db 470 IGGILTPKTKGLELDGGKAQAAAVGLSGDKLFIADTQRLYSAD-RSAFEGDDPKLKMP 528
 QY 607 QHALDE---HFHGDHOLSEFDDHQLNALVKNNFRQCHACPLG-NDHQFHPGWNLT 661
 Db 529 EQANFQLEGVPLGHNRTVGFINGDDGVVHALIKNRQGETHSHALDEQSSKLSQSGWNLN 588
 QY 662 ALVIDNOLGLHHTNEP---HEILDMHGLSLALQEGKLYHFDQITKGWTAEB-SDKQL 717
 Db 589 ALVLNNRGL--TWPPPTAADRLNDRAGLVGLSEGRQIQRWDATPECKWDAGIKDIDL 646
 QY 718 KKGIDGAAYLLKDEVKRLNINQSTSSIKHGTENVFSLPHVYNKPEPGDALQGLKDDKA 777
 Db 647 QRGADSNAYVLKGGKHALKTAABHPNNAFDRNTALATARTKVMKEKELEGL--DRV 704
 QY 778 -QAMAVIGWVKYLALTEKGDIRSFOIKPGTQQLERPAQTLREGISGBELKDIDHVDHKML 836
 Db 705 IKAFAMWNSKRFVALDD-----QNKLTAKSHOKHPVTLDIPGLEGDIKSLSDKEHNL 757
 QY 837 YALTHEGEVFPHPRAWQ---GAB-SSSWKHALPQSEKLSLDMSHHEHKPIATFDDG 892
 Db 758 HALTSTGGLYCLPKQAWQSTKLGDQLRARWTPVALPGGQ-PVKALFTNDNDVLSAQIEDA 816
 QY 893 SOH-----QLKAGWHAYAPERGFLAVGTSGSTVENRLMQGVKGVIPGSGTLVKLSAQ 949
 Db 817 ECKGLMOLKAGOWQRF---EQRP--VEENGLNDVHSRTGSKNTWRIPKIGITLMDVNT 871
 QY 950 GQMTGAEGRKSSKFSERIRAVAFNPTMTSPRIKNAAYATQHQWQREGKPLXEMQCA 1009
 Db 872 FORSGVEKSKAST-SEFIRANIYKNTAETPRWKNVGDHQRHQGRGLKEVETESM 930
 QY 1010 LKQLDAHNVHRNAPQ-----DLQSKLETL- -LGEHGAELNDMKRPRDELOSATRS 1062
 Db 931 LFKQLEL--IHESGRRPPARGQDLKARITALEAKLGPQGATLVKELETLDLEHNSHTA 988
 QY 1063 VTVLG-----OHQVLSKNGEINSEFPSP-----GKALVQ---SFNVNRSGODLS 1105
 Db 989 LMSIGQSYGAKNKLKQDDGILNHQBEL---AKPSVMQFQKGLADLGTKLNFKSSGHDIV 1045
 QY 1106 KSLQQAHVHATPPSAESKLSMLGHFVSAGVDMSHQGEIPLGRQRPNDKLTALTKSRLIL 1165
 Db 1046 KELQDALTOVAPSAENPTKLLAGTLKHQGLKLSHQKADIPLGQRDASEDHGLSKARLAL 1105
 QY 1166 DTVTIGELHELADKAKLVSDHKPDADQIKQLRQOFTTLREKRYESNPVHYTDMGTTHNK 1225
 Db 1106 DLVTLKSLGALLDQV---EQLPPQSDIEPLQKLATLDRVTVGENPVPKVTMDMGFTDNK 1161
 QY 1226 ALEANYDAKAFINAPKEHGVNLTTRTVLESQGSAGELAKLKNLTLLSLDSG-ESMSFS 1284
 Db 1162 ALESYESVKFTLKSFKKADHAVSVNNRAATGSKDQAEKLAGKFKMLKQLEHGDDEVLQ 1221
 QY 1285 RSYGGGVSTVFTLTKKVPVPIPGAGITLDRAYNLSFSRSTSGGLN---VSFRGRDGGVS 1341
 Db 1222 RSYGVNLTTFFIILADK--ATGLWPTAGATGNRYIILNAERCEGGVTLYLISEGA-GNVS 1278
 QY 1342 GNIMVATGHDVMPYM-----TGKTSAGNASDMLSAKHKTSIDPLRIGAAVSTIGTLQN 1396
 Db 1279 GGF--GAGKDYWPGFDPANNPARSDVGN-----NRTLTNPNRGLVDVTATVAASQRA 1329
 QY 1397 SLKPKLTELDPGFIHGLTHGLTLPABELLQKIEHQMKGSKLTFSDVTSANLDRAGIN 1456
 Db 1330 GVVFNVPDEIDAFVDDLFEQLMPLQVLKAVDYESYARRNFNFDLTAGTADIRAGIN 1389
 QY 1457 LNEED--GSKFN--GVTVARVSAGLSA--SANAAGSRERSTTSQGFSTTSAGNNRPTFL 1509

Db 1390 LTEDRDPNADPNSDSFSAVVRGGAANITVNLMTYTDYSITQKNDKTELKEGKNRREFL 1449
 QY 1510 NGVGAGANLTAALGVAAHSSTHEKPV---GIFPA--FTSTNVSAAL--ALDNRTSQSISL 1562
 Db 1450 NNTVAGQLRAQIGGSHAP-TGTPASAPGPTPASQTAANNLGCALNFSVENRTVKRIKF 1508
 QY 1563 ELKRAEPTVNSDISELTSTLGKHFQKOSATTYMLAALKELDDA-----KPAE---QJHIL 1613
 Db 1509 RYNVAKBITTEGLSKSLGIGLEAFDNTTKAKLAELADPLNARYTGKKPDEVIQAQDGL 1568
 QY 1614 QOHFS-----AKVVGDERYEAVRNKKLVIRQQA-ADSHSMELSGASHSTTYNNLSRNN 1668
 Db 1569 EELFADIPPPKD--NDKQKALRDLKRAAVEHRASANKHSV-MDNARFETSKTNLSGLSS 1625
 QY 1669 DGIV-ELLHKEHFDAAALPASSAKRLGEMNNDPALKDIIKQLQSTPFSASVSMELKGLR 1727
 Db 1626 ESILTKIMSSVRDASAP-GNATRVAEFMRQDPKRLAMLKEMEGSIGTILARVLEPKDSL 1684
 QY 1728 EQTEKAILDGKVGREEVGVLPQDRNNLKVSVSVSQSVKSEGFENTPALLLGTSNAAAMS 1787
 Db 1685 DKIDEGSLNGTMTQSDLSMLEEDNEMRIKRLVVFHTATQAEFTSPPLVSYNGANVS 1744
 QY 1788 MERNIGTINFKYQDQNTPRFTLEGGAQANPOVASALTDLKKEGLEMKS 1838
 Db 1745 VTKTLGRINFGVAGDQDKPIGTTFDGELSRPSASLKERAGDLKKEGFELKS 1795

RESULT 3
 B89921
 Hypochemical protein ebha [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: B89921
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; PMID:21311952; PMID:11418146
 A:Accession: B89921
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-6713 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BA842527.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: ebha

Query Match 2.9%; Score 275.5; DB 2; Length 6713;
 Best Local Similarity 17.8%; Pred. No. 0.00021;
 Matches 407; Conservative 322; Mismatches 904; Indels 651; Gaps 95;

QY 11 KAAVHTAHPVGHGVAL-----QOGSSSSSPQNAASLAAGKNGKMPRIHQFSTA 63
 Db 556 KAAVENALSQVTNAKALNGNHNLEQAKSNANTINGLOHLITTAOKDKLKQ-QVQAQNV 614
 QY 64 ADGISAHHQKKSFSIRGCLGKTKFSRPAQOQPGTHSKGATRLDLLADGETOHEAA 123
 Db 615 A-GVDIV--KSSANTLINGAMGTLRNSI-----QDNTATKNGQNYLD--ATERKNTYNNNA 664
 QY 124 APDAARLTSGGVKRRNMDMAGRPMVKGSGEDKVPQO--QKRHOL---NNFGOMKOTM 178
 Db 665 VDSANGVINA--TSNPNMD-----ANAINQATQVTSKGNALDGNLTQAKQTA 712
 QY 179 LSNMAHPASNAGDRLQHPHPPHPSGSHHEIKPEVPGSTKATTAAHADV-----EIAQ 231
 Db 713 TNAIDATNLNKAK-----DALKAQVTSQARVANVTSIQOTAN 751
 QY 232 E-----DDSEFQQLHQORLARENPFPKLGIVATPISAREPQKLT--AVA 277
 Db 752 ELNTAMGQLQHGIDDENATKQTKYRDAEQSKKTAYDQAVAAAKAILNKQTSNDSKAAV 811

Science 287, 1809-1815, 2000
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
 A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A;Reference number: A81000; MUID:20175755; PMID:10710307
 A;Accession: F81045
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-2514 <TET>
 A;Cross-references: GB:AE002026; GB:AE002098; NID:g7227015; PIDN:AAF42109.1; PID:g722702
 A;Experimental source: serogroup B, strain MC58
 C;Genetics:
 A;Gene: NMB1768

Query Match 2.9%; Score 275; DB 2; Length 2514;
 Best Local Similarity 18.8%; Pred. No. 4.3e-05;
 Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY	7	GTEH-KAAVHTAAHPV-----GHCVALQOQSSS-----SSPQNAAS 43
Db	37	GSAAVKSVPFGTTHAPVCRSIFPSLLGFSCLAVGTANTAFADGIIADKAAKPTQAT 96
QY	44	LAAPGKNGKMRHQPSTADGISAHQKKSFLRGCLGKTKKFSRQPG-----98
Db	97	ILQTGNG---IPQVNIQTPTSAQSVNQYAFDVGNERGAILNN--SRSTQTQLGGMIOG 151
QY	99	-----TTHSK-----GATLRDLARDDGGETOHEAAAPDAARLTR 132
Db	152	NPWLARGEARVVVQINSHSSQWNGYIEVGRRAEVIANPAGIANGGFINASRATL 211
QY	133	SGGVKRRNMDMAGPMVKG-----GSGED-----KVPTQQRHQLNFGQMRQTMLS 180
Db	212	TTGQPYQAGLSGFKPIQGNVVIAGHGLDARDTDFILSVHSKIDAPVWGQVRVAVG 271
QY	181	KMAHPASANAGRLQHPHPIGSHHEIKERPEVSGTSKATTAHADRVHIAQEDDSEFQQ 240
Db	272	QNDVVATGNA---HSP-----ILNNAANTSNNTANNTHPIFAIDTG-----312
QY	241	LHQRLARERNPFPKLGIVATPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
Db	313	-----KLG-----GMVANKITILISTA-----EKGIRNQQLFAS 342
QY	301	GAGVTPAVTLDKGLQAPNDPNAITLLKQTLGKDTQHYLAHAS--SDGSQHLLLDN 358
Db	343	SGN-----VAIDANGELVNSGTMAAN-----AKDONTAEHKVNIRSQG-----VEN 385
QY	359	KG-----HLFDIKSTATSYS-----VLNNSHPGEIKGLAQAGTGSVSDGSKGI 404
Db	386	SGTAVSQOQTQIHSQSIQNTGTLSSGELIHN-----GSLKNETSCTI-----EAARL 435
QY	405	SLGSGTQSHNKTWLSQPGEAHRSLLTGIWQHPAGARPQGESIRLHDDKIHILPELGVW 464
Db	436	AIDTDT-LNNQKLSQTG-----SQKLHI-----458
QY	465	QSADKDTSHLSQROADKGLYALKDNRTLQNLSD-----NKSEKLVDKIKSYSDVQOQ 518
Db	459	-----DAQKM-----DNGRMGLQDTAPTASNGSNQGN-----SYNASEFSS 498
QY	519	VAILTDPGRHKMSIMPSLDAPESHSISLHPADAHQGLHGKSELEAQSVASHGRILV 578
Db	499	TTTTPTATGTGTATVSISNITAPT-----FAD-----GTIRTHGALDNGSGSIANGQTD 547
QY	579	VADSEKLSAALPKQGDGNELKMKAMPQHALDEHFGHDHQLSGFFDHGQLNALVKN- 637
Db	548	VSAAQOG-LANNAG---QIDIHQLNAKG---SAFDNNG--TIIISDAVHIQAGSLNNQNGNI 598
QY	638	NFQQOACPLGNDHQHPGWNLTALVID-NQLGLHHTNPEPEHILDMHGLGSLALQEGK 696
Db	599	TTTQQ-----LEIETDQLDNAHGKLLSAEIALDVAVSGSLNNQNGE 638
QY	697	LHYFDQLTKGWTGAESDCKOLKKGDLGAAYLLKDGVEKVRININQSTSSIKHGTENFSLP 756
Db	639	IATNQQL-----IHDGQOQSTAVIDNTNGTIQSGRDVAIQAK 675

QY	757	HVANKPEPGDALOGLNK-----DDKAQAMAVIGVKNKYLTALTEKGDIRSFOIKPGTQOLE 810
Db	676	SLSN-----NGTLAANKLDIALQDDFYVERNI VAGNE-LSLSTRGSLKUNSHLTQAGRRIR 730
QY	811	RPAQTLISREG-----ISGELDIHVHDKONLYALTTHEGEVFHQPREAWQNGAESSESMHKL 866
Db	731	IKANNLDNAAQNIQSGGTTDGTQHN-----LTNRGLIDGQ-----767
QY	867	LPOSESKLSLDMSH-----EHPFIATFEDGSOHLKAGGWHAVAAERGPPLAYGTS 918
Db	768	-----QTKIQAGQMNIGTIGRIYGNIAAATRLDNQDENGTA--ATAARENLMGLIG-- 819
QY	919	GSQTVFNRLMQGVKVI PGSLTIVKLSAQGTGAGTGAEGRKVSSKFSERIPAYAFNPMS 978
Db	820	-----QLNRENLSIYSGNDMAVGALDNTNGQATKAQR-----853
QY	979	TPRIKAAAYATQGMQGREGLKPLYEMOQALIKQL-----DAHVRHNAP 1024
Db	854	-----IHNAGATIEAAGKMRILGVEKLHNTNHLKTLQVETGREHIVDYEAFGREHLLREGT 909
QY	1025	QPDL-----QSKLETDLGHE-----GAEILLDMK 1049
Db	910	QHELGSVYNDESDDLRTPDGAAHENMHKYDEYKVTQKTQVTPAPAKIISGNDLTDGK 969
QY	1050	R-FRDELEQSATRSVTVLGGHQ-----VLKSGEINSEFKPS-PGKALVQSFN 1096
Db	970	EVFNTDSQIAGGNLI VQTEKDXLHNEQTFGKVKVFBENGKLIHSTYREKHKGR--DSG 1026
QY	1097	VNRSGDLSKLSQAQVHATPPSAESKLQSMGLHFVSAGVDMSHQGEIPLQRQDPNDKT 1156
Db	1027	HSEQNYTLPEBITRNI SLGSPAYESHKALSHHAPSQGTETLPQSG-----ISLPTSNSFT 1083
QY	1157	ALTKSRL-ILDTVTITIGELHELADKAK-----LVSDHKPDA--DOI KOLROQDFTLRKR 1207
Db	1084	PLPSSLYIINPVNKGVLVETDPRFANYRQWLGSDYMLDSLKLPNNLHKLGLDGYEQR 1143
QY	1208	YESNPVHYTMGTFTNKALANYDAVAFINAPKEHGVNLTITRTVLESGSSEALAKK 1267
Db	1144	LINEQIAELT--GHRRLDGYQNDDEQFKALMDNGATAARSNNLSVGIAL-----SAEQVAQ 1197
QY	1268	LKNTLLSLDSGESFSRSYGGVSTVFVPTLSKKVPVPVPIPGAGITLDRAYNLSFRSTS 1327
Db	1198	LTSDIVMLVQKE-----VKLPDGGTQTVLPQVYVVRKNGDIDGKAL-----LSGSNTQ 1247
QY	1328	GGLNVSFORDGVSQNI MVATGHDVMPYMTGK--KTSA-----GNASDWLSAKHKIS 1377
Db	1248	INVSGLKNSGTIAGRNALLINTDILNIGGRIHAKSAVATATQDINNIGGMLSAEQTL 1307
QY	1378	PDLRTGAAY-SGTLOGTQNSLKFKLTEDELPG-FIHGLTHGTLT-----PAELLQKGIE 1430
Db	1308	--LNAGNNINSQSTTASSONTQSSSTYLDRMAGIYITGKEGVLAQAQKDINITIAQGIS 1365
QY	1431	HQMKOGSKLTFPSVDTSANLDEL-----RAGINLNEGSKPENGVTARVSAGLSASA-----1479
Db	1366	NQSEQG-QTRLQAGRDINDLTQVTSKHQTHFDADNHVIRGSTNEVSGSIQIKGDTVLLS 1424
QY	1480	--NLAAGSRERTTSCQFSGTTSASNNRPTFLNGVAGANLTAALGVAHSSTHEGKPVGI 1537
Db	1425	GNNLAKAAEVSSANGTL--AVSAKND--IN-ISAGINTT--HVDDASKHTGRSGG--1473
QY	1538	FFAFTSTNVSAALALNRTSQSISILELKRABPVTSNDISELTSTL-----CKHFK 1587
Db	1474	---GNKLVTIDKAAQSHHETAQSTTFEGKQVVLQAGNDANILGSNVISDNGTIQAGNHVR 1530
QY	1588	DSATTKMLAALKELDDAKPAEQHLHQHFSAKDVVGVDEREYAVRNKLKLVIRQQAAASH 1647
Db	1531	IGTT-----OTQOSEYVTHQTKSGLMSAGIG-----FTI---GSKIN 1565
QY	1648	SMELGSASHSTTVNNLSRINNDQIVELLHKHDAALPASSAKRELGEMM-----NNDPAUK 1702
Db	1566	TQENQSQSNEHTGSTVSLKGDTTI-VAGKHYE-----QIGSTVSSPEGNNTIYAQ 1615
QY	1703	DIITKQLQSTPFSASVSMELKGLREQTEKAILDGKVGREVG-----LFDQRNLRV 1756

```
Db      1616 SIDIQAANHLNSNTQTYEQKGLTVARSPVTD--LAQQAIAVAQSSKQVQSKND-RV 1672
QY      1757 KSVSVSQS--VSKSGFNTPALLLGTSNSAAMERNIGTINFKYQDQNTPRRT-LEG 1813
Db      1673 NAMAANAAGWQAYGTQKGAQNLANGTNAKVQV-----ISITVGEQCN--RQYTVQA 1723
QY      1814 GIAQANPOVASALTDLKKEGLEMK 1838
Db      1724 NQAQASQIQAGGKTTLIATGAEQS 1748

RESULT 5
A35938
Profilaggrin - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 14-Dec-1990 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C:Accession: A35938
R:Gan, S.Q.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.
Biochemistry 29, 9432-9440, 1990
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.
A:Reference number: A35938; PMID:91064347; PMID:2248957
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-2248 <GAN>
A:Cross-references: GB:J02929
C:Genetics:
A:Gene: GDB:FLG
A:Cross-references: GDB:I19912; OMIM:135940
A:Map position: lq21-lq21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: Ep hand; epidermis; polymorphism; tandem repeat
F:246-565/Region: filaggrin repeat
F:570-893/Region: filaggrin repeat
F:1074-1397/Region: filaggrin repeat
F:1573-1896/Region: filaggrin repeat

Query Match      2.7%; Score 258.5; DB 2; Length 2248;
Best Local Similarity 18.4%; Pred. No. 0.00024;
Matches 381; Conservative 249; Mismatches 815; Indels 629; Gaps 83;

QY      7 GTEHKAHVTAHNPFVGHVALQ-----QGSSSSSPQNAASLAAGKMGKMPR 56
Db      487 GSGHQQADSSRHSGIGHGQASSAVRDSGHRGYSQASQOEGHSENSDQSVSGQAR 546
QY      57 IHQPSTAADGISAAHQKKSFLRGLGKTKKFSRQAPQGPQGTTHSKGATLLDLDG 116
Db      547 SHQQS-----HQES-----TRGQSRGSRGSR-----GSZLYQVSTHQ 580
QY      117 ETOHEAAAPPAALTRSGGVKRRNMDMAGRPVMYKGGSDKVPQOQRHQLNPFQMQR 176
Db      581 ESAHGRSAPSTR-----RQSHHDQA-RDSSRHASQEGQ--D 616
QY      177 TMLSKWAHPASANAG-----DLQSPHPIPGSH-----ELKEPVGSTKAT 221
Db      617 TI---RGHPSSRGQSGHYEQSVDRSGH-----GSHSHTTSQGRSDASHGTSGRS 668
QY      222 AHADRVIAQEDDSDPFQQLHQQLARERNPPQPKLGVATPISARFQPKLTAVAESVL 281
Db      669 ASRQTRDEQSGDGRHSGSHHQAESTQADS-SRHQVQVQSGAGSR-----TSRN 718
QY      282 EGTDTTQ---SPLKPSMLKGSAGVT-----PLAVTLDKGLQAPDN 322
Db      719 QGSSVSQDRDEQSGEDSERHSGSASNRHSGSAEQSRDGRHFGSHDEDRAGHQADS 778
QY      323 PPAINTLLKQTLCK-----DTQVLAHHAGSDGSHLLLDNKGHLFDIKST 368
Db      779 SSQSGTRHTQTSSRQAASSQEQARAGDRHSGHQQADSRH-----SGIGRG 829
QY      369 ATSYSVLHNSHPGEIKGLAQAGTGSVVDGSGKTS-----LGSQTQSHNKTMLSQFGEA 424
Db      830 QASSAVRDRGHRGS-RGSAQSDQEGH-SEDSQSVSAQRAQSHGQSHQSHQEST----- 880
```

```
QY      425 HRSLLTGTIWOHPAGAARPOGES-----IRLHDDKIHLHPGLGWQSAKDQTHSQL 475
Db      881 -----RGRSQSGRSGSYQVSTHQES-----ESAGRTTRST 915
QY      476 SRQADGKLYALKONRTLQNLSDNKSSEKLVKIKSYSDVORGQVAILTDTQGRHKMSIMP 535
Db      916 GRRQGSHEHQARDS-----SRHSASQSGQDTIRAHPCGSRG-----GRQSHHEQ 960
QY      536 SLDASPESHISLSLHPADAHQGLLHGKSELAEQAQSVASHGRILVADSEKGFSAALPKQG 595
Db      961 SVDRS-----GHSGSHSHTTSQGRSDA-SHG-----QSG 989
QY      596 DGNELKMKAMPQHALDEHFHDHQISGFFHD--DHGQLNALVKNNFRQHQHACPLGNDHQF 653
Db      990 SGSH-----HQQADSSRHSGIGHGQASSAVRDSGHRGSSGQASDSEG 1033
QY      654 HPGWNLTDAVLIDNLQGLHTNPEPHEIDMLDGHGLSLALQEGKLY-----FDQ----- 702
Db      1034 HSEDSDTQSVSAQAGPHQ---QSHQESTRGRSAGRSRSGSFLYQVSTHQESAHGR 1090
QY      703 ---LTKGWTGA-----ESDCKQLKGLDGAAYLLKDGKVKRLIN----- 739
Db      1091 ARTSTRGQSHHEQADSSRHSTSQEGQDTIRHPGSPSGRHHGSHYEQSVNSTGSHGS 1150
QY      740 -----QSTSSIKKHGTENVFSLP-HVRNKPDPGDALQGLNK-----DDKAQANAV 782
Db      1151 HHSHTTSQGRSDASHGTSGRSASRETHNEQSGDGRHSGSRHQEASSWADSSGHSQAG 1210
QY      783 IGNNKLYALTEKGDIRSQIKPGTQQLERPAQTLSREGISGELKDIIH--VDHKQNYALT 840
Db      1211 QGSSG-SRTSRNQGSSF-----SQSDSQSGSEDSERRSGSASRNHRGSAREQSRDGR 1264
QY      841 HEGEVFHQPREAWQNGAESWHKALPQESKLSLDMSEHHPKPIATFEDGSOHLKA- 899
Db      1265 HPGS--HHEDRAGHGHSDSRQSGTRHTQTSSRRQAASSQEQARAGDRHSGHQQAD 1323
QY      900 GGHAYAAAPBPGPLAV-----GTSGSQTVFNR-----LMQVYKGVIPGS-----GL 941
Db      1324 SSRHSGIGRGOASTAVRDSGHRGSRGQASDNEGHSDESDTQSVAGORAGSHHESQES 1383
QY      942 TVKLQAQGTGMTGAEGKVKSKFSERIRAYAFNPTMTSPRPIKNAAYATQHWGQREGLK 1001
Db      1384 TRGQSRSTGRSGSFLYQVST-----HEQSESSHGWSGTS--- 1418
QY      1002 PLYEMOGALIKOLDANVRHNAQPDLOSKLETILDL-----GEGALINDMKRPRDELE 1056
Db      1419 -TRGQSGHHEQAO--DSSRHSAQ-----EGQDTIHGHPGGHSA-----DSSR 1459
QY      1057 QSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFVNRS---GQDLKSL----- 1108
Db      1460 QSGTRHTESSRGOAA--SSHEQARSSAGERHGHGSHHQSSADSRHAGIGHQQAASAVRDSG 1518
QY      1109 -----QAVHATPPSAESKLQSMGLHFVSAGVDMSHQKEIPLGRQRPENDKTALTCSR 1162
Db      1519 HRGYRGOATDSEGHSESDTQSVSAQAGPHQQAQ---ESARGQSGESSG-----RSG 1571
QY      1163 LILDTVTIGELHELADKAKLVS-----DHKPDADQIKQLRQFDTILREKRYSNPV 1213
Db      1572 SFLYQVSTHQESTHGQSVFPGTGGQGHSHDQAQDSRRHSAQEQGDTIRHPGSSR-- 1629
QY      1214 KHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQGSALACKNLTLL 1273
Db      1630 -----GGRHSGHYEQSVDR-----SGHSGSHSHTTSQGRSDASHGSGSRASQTRN 1678
QY      1274 SLDGSEMSFERSY-----GGSVTVFVPTLSKKVPVPIPGAGITLD- 1316
Db      1679 EQSGDGRSHSGSHHHEASTQAESRHSQAGQCSA--GPRTSR-----QGSSVSQDS 1730
QY      1317 -----RAYNLSPFSTGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTSAGNAS 1367
Db      1731 DSQGHSEDSERWSGSRNHRGSAQEQSRDGS-----RPTSHH-----EDRAGHRHSA-ESS 1782
```

QY 1368 DWLSAKHKISPDURIGAAVSTGLQTLNLSKFLKTEDELPGFTH-----GLTHG 1417
 Db 1783 DQSTHH--AENSNGGAASHEQA-----RSSAGRHGSHHQASDSSRHSIGHG 1832
 QY 1418 TLTPAELLQKIEHQMKQSKLTFSVDTNSANLDRAGINLNEDGSKPVGTVIARVAGLSA 1477
 Db 1833 QASSA---VRDSGRHSGSGQASDSHGSESDTQ-----SV 1866
 QY 1478 SANLAAGSRSTSTGQFGSTTSASNNRPTFLNGVAGANLTAALGVA----- 1525
 Db 1867 SAHQAGSHQOHSQESTRGRSQSGSGSGFLYQVSTHEQSESAGHGRAGPSTGGRQSGSRH 1926
 QY 1526 ---HSTHSGKPV-----GIPPAFTSTNVSAAL 1550
 Db 1927 EQARDSRHSASQGDQTIHPSGRRCRGSGSHEQSVDRSGSHSHSTTQGRSDA 1986
 QY 1551 ALDNRTQSISLEIKRAEPVTSNDISELTSTLGHFKDSATTKMLAALKELDDAKPAQBL 1610
 Db 1987 SHGQSGRSASRETRNEB--QSGDGRHS---GSRHHEASTQADSSRHSQSGQGESAGSR 2041
 QY 1611 HILOQHSFAKDVGDERY-----EAVRNMLKLVIRQQAADSHSMELGSAHSHTTYNN 1662
 Db 2042 RSRQSGSVQSDSPSEAYPEDSERRSASRNHHGSSREQSRDG--SRHPGS--SHRDTASH 2099
 QY 1663 L--SRINNDGIVELLHKHFDALPASSAKRLGEMNNDPALDKIILQLQSTPFGSASVSM 1720
 Db 2100 VQSGPVQSDSTAKEHGIF--SLSQSDSAYHSG-----IQSGSPHSSSYHY 2145
 QY 1721 ELKDLRQTEKAIL-----DKVGREVEGVLFQDRNNLRVKS 1758
 Db 2146 QSEGTQRKGQGLVWRHSGSYGADYDYGESGRHSQGSVSYNSPNVLFKERSDICKAS 2205
 QY 1759 V-----SVSQSVSKSEGFN 1772
 Db 2206 AFGKDHPRYATYINKPGLCHGSHSDISKQLGFS 2239

RESULT 6
 B82519
 hemagglutinin-like secreted protein XF2775 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: B82519
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; PMID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: B82519
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3455 <SIM>
 A:Cross-References: GB:AE004082; GB:AE003849; NID:g9108003; PIDN:AAF85560.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canaro, L.E.A.; Carraro, D.M.; Carrer, H
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins, E
 chado, M.A.; Martins, E.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2775

Query Match 2.7%; Score 255.5; DB 2; Length 3455;
 Best Local Similarity 18.7%; Pred. No. 0.00069;

Matches 414; Conservative 285; Mismatches 831; Indels 679; Gaps 98;
 QY 3 LKSLGTEHKAATTAHNVPVGHGVALQOQSSSSSPQAAASLAAGKNGRMKPRIHOPST 62
 Db 1374 IDNLTGRLVGDHIAL-----HAQTLNTRDETSOGHTAATIAARQRLDICADTLRNTAN 1428
 QY 63 A---ADGISAHOOOKSFSLRGCL--GTTKF--SRSAPOQPGTTHSKGATLRLDALLARD 114
 Db 1429 AMILSDGDAAI-----CATLDNALHATGTATLLDNRSATIDITGLTNITTTTLLANNI--RD 1481
 QY 115 DGETQHAAAPDA---ARLFRSGGVKERNMDDMAGRPMVKGSGEDKVPTQCKRHQLNNF 171
 Db 1482 N---VHTAHAPDVVTEARMEOPHW--RKNQPN-----GGSNFRFTSYDAHDI--- 1525
 QY 172 GOMRQTWLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSSTK-----A 219
 Db 1526 -----YYLNPAIIKDD-----PYITPDQSQIHRAIVRLTPTPOTSAFYVARGGLYA 1570
 QY 220 TTAHADRVETIAQEDDDSEFQOLHOORLARENPQPQPKLGVAATPISARFOPKLTAVAES 279
 Db 1571 SOAERRRMDLTARTGDSVLYYTDR-----QDKQPNPDHVA---ATNHS 1613
 QY 280 VLEGDTITOSPDKQSMKLGSGAGVTPLAVTLDKGKLQAPDNPAALNTLLKQLGKQDTQ 339
 Db 1614 AFIGLDTPOQNERFQT-----VPIITYAPGDDRLTYDSNYG-----TCTDCV 1655
 QY 340 HYLAAHASSDGSQHLNKNKGHLFDIKSTATSYVLHNSHFGEIKGLAQAGTSVSDVG 399
 Db 1656 RLVTWHDYTD--PDHTLID-----MHRG--PNDVR-----DN 1683
 QY 400 KSGKISLGSQGTQSHNNKWLSPQGEAHRLLTGTIWOHPAGAARPOGESTRLH-----D 451
 Db 1684 EKYR---DARTTTQDILNPDAGAPALIQTG-----GAMMIQDTDLRNHYADLLAGD 1733
 QY 452 DKIHILHPELVQWQADKQTHSQLSRQADGKLYALKDNRTLQNLNDKSKSEKLVKIKSY 511
 Db 1734 QTVGLPPEHTKEKSDDEHY-----KRVLLIDNRALQ--LS-----RTDTFNI 1776
 QY 512 SVDQRQOVA-----ILDTDPRHKMSIMPSLDASPEHSISLSLHFAHADHQLHGH 561
 Db 1777 STTYRGKUSEPWSNESRTPPTQIGRITSGHQHIAAQLNNVTDSHTAPEIQLHTVN 1836
 QY 562 KSELBAQSAVISHGRVLVADSEKGLFSAAPKQGGNKLKMKAMPQHALDEHFGHDHQS 621
 Db 1837 PS---QTLSVNVGVITVTDTSPLSLHTVSLADNPGSAGQELTYIPDQSITTP----- 1885
 QY 622 GFPHDDHQLNALVK-----NN-----PRQAHACPLGNHQHFGHPGN 658
 Db 1886 -----NAPIRDPAAPPATVPTGPTLTPNNSLFTLHPDAATLITDPTFLGRP 1935
 QY 659 LTDALVIDNQLGHTNPEPHEILDGMHGLSLALQEGKLHYFDQLT--KGWTGAESDCQ 716
 Db 1936 YTSADTQLHALGDHT--LHKRLGDGYVEQRLIRE-----QLAQLTGRRLDGYTDDQ 1988
 QY 717 LKKGLDGAAYLKQGEVK-----RLNINQSTSSI-----XGCTENVPSLPHVRN 760
 Db 1989 YRALLDAGVTVAQOHLRPGIALSADLAQLTSDIVWLVDQVQLPDGTTTRALVPRLYL 2048
 QY 761 KPFGDALQGLNKDDKAQAMAVIGNVKYALALTEKGDIRSFOIKPGTQQLERPACTLSREG 820
 Db 2049 RPTGTD-----LTPDQALLAAASTTINAH--TFNTGTIDARHL-----IDINHTWDOQG 2097
 QY 821 -----ISGELKDIHVHDKQNLVYALTH--EGEVFHPQ--REAWQNGAESSSMHKL- 865
 Db 2098 GRLTADAIHTHTGDTFTLGGQFKARGYLKVHAQGNFLASSTLRDATTCTGRHHSVTELD 2157
 QY 866 -----ALPQSESKLSIDMSHEHKPIATFEDGSOHLKAGWHAYAPAPRGPLAYGT- 917
 Db 2158 QQAGFTVTGPGAYLGL-STDQAMTHNGVAINTGTD-----GYTSLNA--TGPLHLGTL 2208
 QY 918 -----SGSQTVFNLMOGVKVPKPSGLTVKLSA----- 947
 Db 2209 TTHRSPTQWDPNRSRHSRIDTEYGTSGITG--NGDIQLNSGQDINLRAATLHSTQGITAL 2267

QY 1800 GQDQWTPR-----RFTL-EGGIAQNPQVASALTDLKKEGLE 1835
 Db 3231 GGDQHAQQLGSDLGSVIADVGFAALAAAGTAKAAEILGEAGISLSKDVLE 3279
 RESULT 7
 E82599 hemagglutinin-like secreted protein XF2196 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: E82599
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: E82599
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3442 <SIM>
 A:Cross-references: GB:AE004032; GB:AR003849; NID:99107324; PIDN:AAF84995.1; GSPDB:GN00
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, P.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, A.; Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 Submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.B.; Kuramae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F.
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; ;
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2196
 Query Match 2.6%; Score 247; DB 2; Length 3442;
 Best Local Similarity 18.7%; Pred. No. 0.0018;
 Matches 414; Conservative 292; Mismatches 821; Indels 692; Gaps 100;
 QY 3 LKSLGTEHKAHVTAHNPVGHGVALQGGSSSPQNAASLAAGKRNKMPRIHQPST 62
 Db 1374 IDNLGTGRLYGDHIAL-----HAQTLNTRDETSDGHTHAATIAARQLDLCADTLRNTAN 1428
 QY 63 A---ADGISAAHQQKSFSLRGCL---GTYKF--SRSAPOGPGGTHSKGATLRDLIARD 114
 Db 1429 AMILSGDDAAI-----GATLDNALHATGTATLDNRSATDITGTLLNTITTTLLNNI--RD 1481
 QY 115 DGETOHEAAAPDA---ARLTRSGGVKRRNMDMAGRPWVGKGGGDEKVPQKHFQLNNF 171
 Db 1482 N---VHIAHAPDVVTEARMEQPHW--RKQPN-----GGSGNFRFTSNYDAHD!--- 1525
 QY 172 GQMRQTMLSKMAHPASANAGDRLOHSPPHIPGSHHEIKEBPVGSTK-----A 219
 Db 1526 -----YILNPADIKDD-----PYITPDGQIHRVAIVRLTPQTSAYFYARGGLYA 1570
 QY 220 TTAHADRVELAQEDDDSEFQQLHQBLARERNPPQPKLGIVATPISARFPQKLTAVAES 279
 Db 1571 SQAERRRMOLTARTGDSVLYYYTDR-----ODKQPNPDHVA--ATNHS 1613
 QY 280 VLEGTTTQSPKPSQMLKSGAGVTPLAVTLDKGLQ LAPDPNPALNTLLKOTLGDQTK 339
 Db 1614 AFIGLDTPOONERFQI-----VFYITAPGDDRLITYDSNYG-----TCTDDCV 1655
 QY 340 HYLAAHSSDGSQHLLLDKNKHLFDIKSTATSYVLNHSHPGIEIKGLKAQAGTGSVDG 399
 Db 1656 RLVTWHDYTD-PDHTLID-----MHG-PNDVR-----DN 1683
 QY 400 KSGKTSLSGSGTOSHNTKMTLSQGRAPHSITLTGTHOHCACAAPOGCESTH----- 451

Db 1684 EKYR-----DATRTTQDILNPDGAPALIQTG-----GAMMIQTOTLRNHYADLLAGGD 1733
Qy 452 DKHIIHPELCWQWQADKTHSOLSRQADCKLYALKDNRTLQNLSDNKSSEKLVKIKSY 511
Db 1734 QTVGLPPEHPTKESDDEHKY-----KRVLLIDNRLAQ-LS-----RTDTFHHI 1776
Qy 512 SVDPQGOVA-----LITDTPEGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHG 561
Db 1777 STTYRGKVPWNESESTTTTQIGRITSGGHCHIAAQTLNNVDTSHAPEPIQLHTYN 1836
Qy 562 KSELEAQSWAISHLRVADSEKLFSAAIKQSGDNELKMKAMPQHALDEHFGHDHQS 621
Db 1837 PS---TQTLVWNGVITVDTSPSLHTVSLADNGSPSAQELTYIPDQSITTP-----1885
Qy 622 GFPHDDHGQNALVK-----NN-----NQOHACPLGNDHQFHPGMN 658
Db 1886 ---NAPTRDPAAPAVTVTPTGPTLFPNLSLFTLHPDAATLITDPRFTTGRP 1935
Qy 659 LTDALVIDNOLGLHHTNPPEHIELDMGHLSALQEGKLYHFDOLT--KQWTAESDCQ 716
Db 1936 YTSADTQIHALGDHDT--LHKRLGDGYEQLIRE---QLAQLTGRRRLDGYTDDDDQ 1988
Qy 717 LKKGLDGAAYLLKDGVEK-----RLMNGSTSI-----KHGTENVFSLPHVRN 760
Db 1989 YRALLDAGVTVAQHQHLPGLRPGIALSADQLAQLTSDIVMLVQDVLPGDTTTRALVPRLYL 2048
Qy 761 KPEPGDALQGNKDKQAQMAVIGVKNYLALTEKGDIRSFOIKPGTQQLERPAOTLSREG 820
Db 2049 RPRTGD---LTPDGAALLAASTTINAH-TFTNTGTTIDARHL-----IDINAHMDOQG 2097
Qy 821 -----ISGELKDIIHVDHKQNYALPH-BGEVFPQP--REAWONGAESSSHKL- 865
Db 2098 GRUTADAIHHTTGDFTLTGQFKARGYLKVHAQGNFLASLTLRDAATTQGRHHSVTELD 2157
Qy 866 -----ALPOSEKLSLDMSEHKPIATEDSGHOLKAGWHAYAAPERGPVAVGT- 917
Db 2158 QOAGFTVTGPGAYLGL-STDQAMTHNGVAINNTGTD-----GYTSINA--TGFLHLGTL 2208
Qy 918 -----SGSOTVFNLMQGVKVIKPGSLTVKLSA-----947
Db 2209 TTHRSDDTTQMDPRNSHRIDTEYGTSTG-NGDIQLNSGQDINLRATLHSTQCTITAL 2267
Qy 948 QTGGMT-----GAEGRKVSXKFSERIRAYAPNPTWSTPRPKNAAY 988
Db 2268 ATGNVTIHTGDTIQYTSQDSHTKRSGLNSRTTTHAQOQQTQALGSTLSADKVPVKGN 2327
Qy 989 ATQHWQMGREGKLPYEMQGALIKOLDANVR-----HNAPODLOSKELETDLGHHGAE 1044
Db 2328 ITVTGSHVVSDAGTYMQAEHDLTLQAATHTTQSTYSHTTKQRL-----IRNGGASL 2379
Qy 1045 -LNDMKRFRDEQSATRSVTVLGOHQ-----VLKXNGEINSEFPKPS 1087
Db 2380 TLGQSQRTDSTTTATTTGSLIGATNGVLLAGHYQQIGSDVLSLPHGDIDIIHAKVD 2439
Qy 1088 GKALVQSNVNRSGDL-----SKSLQAQVHAHTPPSAESKLSQMLGH 1129
Db 2440 ---IIQAHTSQTQHTATROSLTVALSTPLIAGTAQQAQQAARSQDPRLQALAGL 2496
Qy 1130 FVSAG-----VMSHQ-----KGEIPLGRDNDKNTALTKSLIILDTVTIGELHELA 1177
Db 2497 TTALGAKNTIDAVRQDPRALGGLNASLTVGRSTHSDSTTTTSTAAGSNVTAGGNVHISA 2556
Qy 1178 D-----KAKLVSDHKPDAD-QIKQLRQOPDTLREKRYENPV-----1213
Db 2557 TGDGTASTLTIQSGDVRGDTMYLKADGDIALAAQNTVNTQDRNGRSAGVGVAVNLGS 2616
Qy 1214 -----KHVTDMGFTHNKALEANTYDAVKAFINAFKKEHHGVNLTTRTVL 1256
Db 2617 GTSAGLTAHASTTSGSGHSTDLTWSNVHVGGNLLAIDAGDGLMK---GAIGTAKHVI 2673
Qy 1257 ESQGSAAELAKLNTLLSDSGESMSFSRSYGGVSTVFVPTLSKKVPVPIPGAGITLD 1316
Db 2674 ----ADIAGNLRIESLQ-DTHQYRSTDRSLGSLTA-----GAGFSGS 2711

Qy 1317 RAYN-----LSFRTSGGLNVSFGRDGGVSGNINMVAIGHDVMFVMTGKTSAGNASD 1368
Db 2712 ANLNHQITIRSDYASVTEQSG---LFTGDGGYQLTVGGQT-HLIGGAITSNIAHNLGN 2766
Qy 1369 WLSAKHIKISPLRIGRAVSGTL-QGTILQNSLKFILTEDELPGFI--HGLTHGTLTPAE 1425
Db 2767 SLD-----TGTLQNIENHANTATQVNLGGYSGRNGTGT-----2804
Qy 1426 QKGIHQMKQSGKLTFSVDTSANLDRAGINLNEDGSKPNGVTARVS-----AGLSASA 1479
Db 2805 -----DQGHATATQVEGTLPSHNLGSL---ASPPAMTARDSSHSTYSIGISOGA 2853
Qy 1480 NLAAGSRERTTSGQFGSTTSASNNRP-----2856
Db 2854 LTRDDDTAQHALTGHTAAETIATLNRLDITATSNALTPIDFQRINAGFDIVSGLQRE 2913
Qy 1507 --TFLNVGVGAGANL-----TAALGVAH-----SSTHEGKEV--GIF 1538
Db 2914 TGTFINNRAAEADLTKRQATAADHAAHDPNGFNDQORHARERQAIATNEAHAIKDAMG 2973
Qy 1539 PAFSTTVNSAALDNRTSOSISLELKRABPVTNSNDISELTSTLTKHF-----KDSAT 1591
Db 2974 PGTYRQITTTALAAG--ASGNVS-----AASNVSAASSDLAKHVIWVYVQOQATA 3023
Qy 1592 TKMLAALKELDADKAPAE-QLHIL-----QOHFSA-----KD 1621
Db 3024 IGHVATGQLTEGSPHLAALHALLACAGAAASQOHCSSGAQAAASSVLTLGFSDDPRPD 3083
Qy 1622 VUGDERYEAVRNKLVIRQOADSHELSGASHSTYNNLSRINNDGI-----VELLH 1676
Db 3084 TTQDR--EAKRNLTTSTVTGASTGNT-DAATAHAA-----IAAVDNNWLAQKQVQVQMLN 3136
Qy 1677 KHPDAALPASSAKRLGEMMNDPALKDIIKQLQSTPFSASVSMELKDLREQTEKALD 1736
Db 3137 EEFERA---TEKEKRL-----EEKVRAKWRB-IDARQDKLTVD 3172
Qy 1737 GKV-GREEVGLVFQDRNNLR---VKSVSVSVOSVSKSEGFNTPALLLGTSNAAKSMERNI 1792
Db 3173 GLLAKGLKESGI--NDINGLEHLFHPVDTVHELKGI--LTHPTLLQLGESAVQELLNKV 3228
Qy 1793 GTIN--FKYGOQNTPRFTLEGG--TAQANPQVASALT-----DLKKEGLE 1835
Db 3229 SRMEALLVGGDOHA-QQFGEDLGSVIADVGVALLAAAGTFRKAAEILGEAGINLSKDVLE 3286

RESULT 8

H81193
hemagglutinin/hemolysin-related protein NMB0493 [imported] - Neisseria meningitidis (str)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81193
R:Testelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gall, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: H81193
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2703 <TST>
A:Cross-references: GB:A8002405; GB:A8002098; NID:g7225708; PIDN:AAF40927.1; PID:g7225711
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0493

Query Match 2.6%; Score 245; DB 2; Length 2703;

Best Local Similarity 19.0%; Pred. No. 0.0015;

Matches 414; Conservative 282; Mismatches 766; Indels 712; Gaps 105;

Qy 10 HKAAYHTAAHNPVGHGVALQQGSSSSSPQNAASLAAREGKNRGMKPRIHQPSTAAD---G 66

Db 573 HDIAVNTQTAKNSGH-LLTGTGKIDNRELHNAAGIAA-----NNLTLIHSGRSLNDXKGN 626
QY 67 ISAAHQKKSFLR-----CCLGTTKFSRPAQOQPGTTHSKGATLRLDARDGE 117
Db 627 IRAHLQDITAGLHAGNILLADSTVTTKNLRITGVSVARLNTGOTLONTGRJEA 686
QY 118 TOHEAAPDAARLTRSG-----VKRRMDMAGRPM-----VKGG-----SGE 156
Db 687 TVN-----IQSOQLTQSGHITATBOLTINSRNVNQKLLSANAQALVSDGLYNQHG 742
QY 157 DKVPTQKRLNFGQWRQMTLSKMAHPASANAGRLQSPHPHPOSHHEIKEEPVGST 216
Db 743 IATNRQJLSIHDKN-----QNTLANNADGTIQSAGN-----VSLQAKSLA 782
QY 217 SKATTAHADVEIAQEDDDSEFQQLHQORLARENPFPKPLGVATPISARFOPKLTA 276
Db 783 NNGTLTAGNKLDIALTDD-----FVERD-----LTAGKQLNS-- 816
QY 277 AESVLEGTDTQSPKFPQSMKGGAGVTPLAUTLDKGLQLADNPALNLLKOTLGK 336
Db 817 IKGLKNTHTLO-----AGHT---LXINAGNI-----DN-----QVTGK 847
QY 337 DTQHYLAHASSDGSQHLLDNKGHLFDIKSTATSYSLHNSHPGEIKGLAQAGTGSV 396
Db 848 -----ITGSGQTDITSEHVDNRG-----LINS-----DGLTHIGAGQTL 882
QY 397 VDGKSGKISLGSQTSQHNKMTLSQGBAHRSLLTGIWOHPAGAAPQGE---SIRLHDDKI 454
Db 883 TNGTGTGI-YGNHIALDAQILLNREETEGSTAG-----AIAARKLDIGAKEIHNQEG 936
QY 455 HILHPE--LGVMSADKTHSQ-----LSRQADGKYALKDNFTLQNLSDNKS 500
Db 937 ALLSSEGIFAVGNRLDQHEHAAGMADTFVNGSAGLEVQGD-----ALMSVRMOMINNHFK 992
QY 501 SE-----KLVDKIKSYSV-----DORQOVALITDTPGRHKMSIMPSLDASPESHISLS-L 549
Db 993 TETYLAKAEQVADYTVILGQNTYVOAGKGLFNSQOQ-KDQTATPHLANGSRIEANQW 1051
QY 550 HFADAH-----QGLLH-----GKSELEAQSAVAISHGRVLVAD-----581
Db 1052 HVRDYHETTYKERIENRPAHITVGGDLTAGSQWNLNKDSRIVVGRIITDDLNQKEITN 1111
QY 582 --SEGKFLSAIIPKQD-----GNELMKAMPQHALDEHFGHDHQISGFHDDHQ 630
Db 1112 QSTGTGRITDAVGTQWDVTKGWSGRKQRERNHTP-----YHDTQI--FTHDFDTP 1165
QY 631 LNALVKN-----NFRQOQHAC-----PLGNDHQHPG--WNLTDAVITDNQGLHHTN 675
Db 1166 VSVIQNAASPFPQPAASAIKLIDGVSTAAVNGQRIHTGNVSVLNNATVLPNSLYTTH 1225
QY 676 PEPHEILDMHGLSLAQEGKLHYFDQLTGKWTGAESDCKQLKKGLDGAAYLLKDGVEKR 735
Db 1226 PD-----NKGW-----LVETDPQFADY--RRWLGSYMLQLQLDNLHLKRLGDGYEQ 1273
QY 736 LAINQSTSSIKHGTENVFSLPHVRNRPEDGALQGLKDDKAQAMAVIGNVKYLALTEKG 795
Db 1274 KLVN-----EQIHQLTGYRR-----LDGYRSDEE-----QPKALMDNG 1306
QY 796 --DIRSFQIKPG--TOQ-----LERPAQTLREGISGELKD IHVDHKQNLXAL 839
Db 1307 LTAAKTFLGTLPGALSABQVARLITSDIVMENGQTVLS-----DGSTQTVLIV--PKVYAL 1359
QY 840 THEGEV-----FHQPREAWNGAESS--WHKLALPQSES-----872
Db 1360 ARKGDLNTSGGLISAEQVLLKLQNLNTSGTITAGRQAVLIQARNINSNGIQADQIGLK 1419
QY 873 KLSLSDMSHEHKPIATFEDGSQHQ-----LKAGWTHAYAAAPER-----G 911
Db 1420 AEKSIINI-----DGGQVOAGRLITTAQONINLNGTTQTSNGERNNGNTAIDRMAG 1468
QY 912 PLAVGTSGSQTVNRLMQGVKGVIPGSGLTVKLSAQTCGTMCAEGRKVSXKFSF---RI 968

Db 1469 INVVG-SHTEQVDRNTSDGI-----LSLHASNDINLNAATVSNQVKGDTQI 1514
QY 969 RAYAFNTMTSPPIKNAAYATQGWQREGKPLIYEMQGALIKQLDAHNVHNAPODPL 1028
Db 1515 TA-ONNLNLGTIRTEHREAYGT-----LDDENHRHRQSTEV 1550
QY 1029 QSKLETLDLGHEGAELL--NDMKRFRDELEQ-----SATRSVT-----VL 1066
Db 1551 GSSIRT-----QNGALLRAGNDLKIROGELEABECKTVLAAGRDVTISEGROITELDTSVS 1606
QY 1067 GQHGGVLKNGEINSEFKPS-----PGKALV--OSFNVN-----RSG 1101
Db 1607 GKSGILSST-KTHDRYRFSHDEAVGSGNIGGKMIVAAGODINVRGSLNLSIDKGLVKAG 1665
QY 1102 QDLKSILQQAHAHTPPSAESKLOSMGLHFYSAGVDMSHQKEIPLGRQR--DPNDKKTALT 1159
Db 1666 HDIDISTAHNRYTCNEYHESKSGVMG--TGGLGFT-----IGNRKTITDDTDRNIV 1715
QY 1160 KSRLLIL-----DVTIGELHELADKAKLVSDHKPDADQIKQLROQFOTLREKRYESFVK 1214
Db 1716 HTGSIIGSLNGDITVTAGNRYRQTGSTVSSPEGRNTVTAKSIDVEF--ANNRATDYAH 1772
QY 1215 HYTDMGFT--HNKALEANDYDAKAFINAFKKEHGVNLTRTVLESQGSABEAKUKLNT 1271
Db 1773 TOEQKGLTVALNPVVOAQNFIQAQNVGSKNKRNVMAAANAAMQ-SYQATQOQOQ 1831
QY 1272 LLSLSDGESMSFSRSGYGVSTVFPVPLSK-----KVPVPVIPCAGITLDRAYNLSP 1323
Db 1832 APSSAGQGNQNSPISVSIITGQKSRNEQKRHYTEAASQIIGKGT-----TL 1884
QY 1324 SRTSGG--LNVSGRGGVSGNIMVATGHDVW--PYMTGKKTSGAGSNADW--LSAKHKI 1376
Db 1885 AATGSGEQSNITGSDVIGHAGTALIADNHRILOSAQKQDGSQSKNKGWAGVAVKI 1944
QY 1377 SPDLRTGAAVSGLT-QGTIONSILKFKLDEDELPGFIHGLTH-----GTLT 1420
Db 1945 GNGIRFCITAGGNIKGKEQGG-----STTHRHVHGSTTKTIRSGGDTTLK 1993
QY 1421 PAELLQKGT-----HOMKO-----GSKLT-----1440
Db 1994 GVOLIKGKIQADTRNLHIESVDTEYQSKQNGNVQVTVGYGFSAGSVRSQKVADHA 2053
QY 1441 -----FSDVTSANLDRAGINLNEBDSKPNGVTVARVSAGLSASANLAAGSR 1486
Db 2054 SVTQSGIYAGDGYQIKVRDNTDLKGGIITSSQSAEDKGNLFQATLTAS-DIQNHSR 2112
QY 1487 ERSTTSQSGS-----TTSASNRPFTFLNGVAG-----ANLTAALGV---AH 1526
Db 2113 YEGRSFGIGSFDLNGWDGTVTDKQGRPTDRISPAAGYSGDGSKNSTTRSGVNTNHI 2172
QY 1527 SSTHEKPVGIFPAFTTNVSAALADNRTSQSISLELKRA--EPVTSNDIS--ELTST 1581
Db 2173 ITDEAGQLARTGETAKETEARIYTGIDTETADQHSGLHKNSPKDAVAKEINLOREVTK 2232
QY 1582 LGKHFKDSATTMLAALKBLDDAKPAEQLHLIQQHFSAKDVWGDVEYEAARNLKLVIQR 1641
Db 2233 FGRN-----AAQAVAAV-----ADKLNGTQSY-----ERYQEARTL--LEAEL 2268
QY 1642 QAADSHSME-----LGSASHSTYNNLSRINN 1568
Db 2269 QNTDSEAKAAAFRASLGQVNAVLAENQSYDTWKEGGIGRSILHGAAGGLTTGSLGILA 2328
QY 1669 DGIIVELLHFKHFAALPASSAKRIGEMMNDPALKIIKOLQSTFPFSSASVSMELKDGLRE 1728
Db 2329 GGGTSLAAYLQKA-----AENLG-----PAGKAVALNAGGAATGYAT-----2366
QY 1729 QTEKAILDGKGRBEEVGLFQDRNN--LRVKSVSVS-----QSVKSEGFNTPALLL 1778
Db 2367 -----CGSGAVVGVANV-DWNNRQLHPKEMALADKYAEALKREVEKREGKI-----2412
QY 1779 GTSNAAAMWERNI 1792
Db 2413 -SQEAMMRIRQI 2425

Db 2080 FKVLDDTQGEARVLDNLKLKELQSNKESVSOMKQD-----EDLERRLEQAEKHL 2135
QY 1587 KDSATTWLAALAKLDDAKPAEQHLILQHFSAKDVVDDEYEAVERNKLKLVIRQQAADS 1646
Db 2136 KEKN-----MQBKDALRREKHL-----BETIGEVTLUNKDKKEVQOQJQENLDS 2182
QY 1647 HSMELG--SASHSTYNNLRINNDGIVELHHEFDALPASPASSAKRILGEMNNDPALKDI 1704
Db 2183 TVTQLAAFTKSMSSLDQDRDV-IDEAKWKEKFSDAIQSKEEBIRLKE--DNCSVLKQD 2239
QY 1705 IKQLOSTPFSASVSM-ELKDLG-REQTEKAILDKGVRE----- 1742
Db 2240 LRQM-----SIHMBELKINTSRLEHDKQIWESKAQTEVQLQCKVCDTLQGENKELLS 2291
QY 1743 -----EVGVLFODRNLNRKVSQSVSKSEGFNTPALILGTSNSAASM 1768
Db 2292 QLEBTRHLYHSSQNELAKLELSKLDQLTDLNSLEKCKEQ-----GNLEGIIRQ 2345
QY 1789 ERNIGTTFNKYGQ---DONTFRPT--LEGGIAQANPQVASALTDLKKEGLEM 1836
Db 2346 EADTONSKFSYEQLETDQASRELTSRLHEEINNKEQKIISLLSG-KEEAIQV 2397

RESULT 10
I52300
gi|152300|human
N;Alternate names: gcp372
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C;Accession: I52300
R;Sonda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A;Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A;Reference number: I52300; MUID:55100974; PMID:7802676
A;Accession: I52300
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3225 <RES>
A;Cross-references: GB:D25542; NID:G662389; PIDN:RAA05025.1; PID:G808869
C;Superfamily: giantin

Query Match 2.6%; Score 243.5; DB 2; Length 3225;
Best Local Similarity 18.6%; Pred. No. 0.0024;
Matches 389; Conservative 311; Mismatches 816; Indels 577; Gaps 92;

QY 23 GHGVALQGGSSSSPQNAALAAEGNRCMKPRIHQ--PSTAADGISAAHQCKKPSLR 80
Db 569 GEGIA-----PIKMKVLEDTGQDFPLMNESSILPAVEKQASTBHQSTSEI- 618
QY 81 GCLGTKKFSRSPQGGPQTHSKGATLRDILLARDGGETQHEAAPDAAR----- 129
Db 619 -----SLNDAGVELS--TKQDGD-XSLSAVPDQCHQDELERLKS 657
QY 130 -----LTRSGVKRRNMDMAGRPVVKGGSGEDKVPQCKRHQNN-----FGQMR 175
Db 658 QILELELNPHKQAEIYEKNLDEKA-----KEISNLNQLIEEFKKNADNNSAFTALSEER 712
QY 176 QTMLSKM-----AHPASANAGDRLOHSPHIPGSHHEIKERPVGSTSK 218
Db 713 DQLLSQVKELSMVELRAQVKQLEMLNLAERQRRLDYES--QTAHNLTLTQIHLSLI 769
QY 219 ATTAHADRVRIQAD--DSEFQQLHQORLARENPFPQPKLV-----ATPISARFP 271
Db 770 EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRLSQSQLOKNESEVLEGAERVVRHISKVEE 829
QY 272 KLTAVAESVLEGGTDTTQSPKLP-----QSMKKGSGAGVTPPLAVILDKKLQALPDN-- 322
Db 830 LSQALSQKELEITKMQQLLEKKRDVETLQOTIEEKDQVQTEISFSMTERRKQVLENEEKS 889
QY 323 -----PPALNTLKTQIGKDTQHYLAHASSDGSQHLLDNKGHLFDIKSTATSYV 374
Db 890 LGVEIKTLKQNLRLRAEAKKEQVEEDNEVSSGLKQ-----YDEMSPAQISK 940

QY 375 LHNSHPGEI-----KGKLAQAGTGSVSDGSGKIS---LGSQTSHNKTMLSOP-- 421
Db 941 EELQHEFDLLKENEQKRKLQALINRKELLQVRSLREBELANLKESKEIPLSETER 1000
QY 422 GEARHSLTGTWHPAGAARPOGESIRLHDQKHILHPELGVMSQADKTHDSQLSROADG 481
Db 1001 GEVEEDKENKEYSEKCVTSKQCEIYI---KQTISEKVEL-QHIRKDLKLEKLAABEQF 1056
QY 482 KLYALKDNRTLN-----LSDNKSSEKLVKIKSYSD--ORQGVAILTDTFGR 528
Db 1057 QALVKQNNQTLQDKTNQIDLLQAEISENQA---IIQKLITSNTDASDGSVALVKET--- 1110
QY 529 HKMSIMESILDAPESHISLSLHFADAHQGLHKGSELEAQSVAISHGLRVVADSEGKLF 588
Db 1111 --VVISPCTGSE-----HWKPELEKILALEKEX---EQLOKKLOE 1148
QY 589 A-----AIPQGD-----GNELKMKAMPOHALDEHFGHDHQSIGFFHDHGOALNALVKN 637
Db 1149 ALTSRKAILKKAQEKERHLREELKQKDDYNRLQEQDEQSKENENIGDQLRQLQIQVRE 1208
QY 638 NF-----RQOHACPLGNDHQHPGWNLTALVIDNQLGLHHTNPEPHEILDHGLGSL 690
Db 1209 SIDGKLPSTQOESC-----SSTPG--LBEPLFKATE--QHTTOP-----VLE----- 1247
QY 691 ALQEGKLYHFDQLTKGWTGAESCKQKGLDGAAY--LLKDGVEVKR---LNINQSTSS 744
Db 1248 -----SNLCPDWPSHSEDSALOGGTSVAQIKQAQLKEIAEKVELELKVSSITSE 1297
QY 745 IKGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAMAVIGVNYKYLATEKGDTRSPQIKP 804
Db 1298 LTKKSEVFLQEQINK-----QGL-----EIES--LKT 1324
QY 805 GTQQLERPAOTLREGISGELKDIHVHDKQNLALYALTHEGEVFPQPRE---AWQNGASSSS 861
Db 1325 VSHEAEVHAESLQCKLESSQLIAGLEHLRELQPKLDELQKLSKKEEDVSYLSGQJSEK 1384
QY 862 WHKLA-----LPOSESKLSLDSMHEHKPIATPEDSGHQKAGGMHAYAAPERGPLAVG 916
Db 1385 EAAALTKIQTETIEQEDLILKALHTQLEWQ--AKEHDERIKQLQVCELCEMKQKPEE---IG 1438
QY 917 TSGSQTVFNRLMOGVKGVIPGSLIVKLSAQTCGTMGAERKVKSSKFSERI---RAYAF 973
Db 1439 EE-----SRAKQOIQKQL-----QAALISRKALKENKSLQBELSLARGTIE 1480
QY 974 NPTMTSTPRPIKNAAYATQHCWQGRE-----GLKPLYEMOGALITKQLDAHNVHRNAPQPD 1027
Db 1481 RLTKSL-----ADVQSVAQNKEDTVLGRLLALQEEERDKLITEMD-----RSLLENQS 1530
QY 1028 LQSKLETDLGERHAGELLNMDKFRDELEQASRSTVTLGQHQGVLSKNGEINSEFKPSP 1087
Db 1531 LSSSCESLKLALSG--LTEDKELVKIEISLKSASKIAESTEWO--EKHKELQKEY---- 1581
QY 1088 GKALVGSFVNRRSGQDLSKLSQAQVHATPPSAESKLSQMSLGHFVSAGVDMSH-----QKG 1142
Db 1582 -EILLOSQ-----ENVSNEAERIQHVVEAVRQEK-QELYGKLSTEAANKKETEKOQLEA 1633
QY 1143 EITPLGQRDPNDKTLTKSLILDTV-----TIGELHELADKAK-----LVSPHKPDADQ 1192
Db 1634 EQEMEEKMKMKRFAKSKQKQKILEEENDRLRAEVHPAGDTAKECMTLSSNASMKEE 1693
QY 1193 IKQLRQOQFDLREK-----RYESNPVHYTDMGFTHNKALEANYADVAKFINAF 1241
Db 1694 LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE---GNVSKQANLEAT----- 1742
QY 1242 KKEHHGVNLTTRTVLESQSAELAKKLNKLTILSDS-----GESMSFSRSYGGGVSTVF-- 1295
Db 1743 EXHDNTNVTBEGTQSIGETE-----EQDSLSMSTRPTCSSESVPSKSAANPAVSKDFSS 1797
QY 1296 -----VPTLSKKVPVPVPIGAGITLDRAYNLFSRTSGG-----LNVSFGRDGV 1340
Db 1798 HDEINNYLQIIDQLKERI-----AGLEBKQKNEFQSTLENEKNTLLSQISTKQDEL 1850

```
QY 1341 -----SGNIMVATGHVMPYMTGKTSAGNASDWLSAKHKISPDLRIGAAVSGTLOQ 1392
Db 1851 KMLQEEVTKMNLNQIQBELSRVTKLKEAABEKD-----DLEE 1890
QY 1393 TLONSILKFKLTDEBELPGFIHGLTHGLTPAELLOKGLHOMKQSGKLTFFSV-----1444
Db 1891 RLMNQLA-----ELNGSI--GNYCQDVYDAQIKNELLESEMKNLKCVCSELEEKQOOLVK 1943
QY 1445 --TSANLDDR-----AGINLNEGSKPNGV-----TARVSAGLS 1476
Db 1944 EKTQVSEIRKEVELEKIQGAQ--KEPGNKSHAKELQELLKEKQEVKQLQKDCIRYQEKIS 2002
QY 1477 ASANLAAGSRERSTSQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSTHEGKPVG 1536
Db 2003 A-----LERTVKALEFVQTESQKDLITKEN-----LAQAVEHRKKAQAEALS 2045
QY 1537 IFPAFTSTNVSALAL--DN-----RTSQSISLEIKRAEPTVTSNDIS--ELSTLKGHF 1586
Db 2046 FKVLDDTQSEAAARVLADNLKLLKELQSNKESVKSQMKQD-----EDLRRLEQAEKHL 2101
QY 1587 KDSATTKMLAALKELDDAKPAEQHLILQHFSAKDVVGDERYEAVRNLLKLVIRQQAADS 1646
Db 2102 KEKN-----MOEKLDALREKHL-----BETIGEIVTLNKDKQKEVQOQLENLDS 2148
QY 1647 HSMELG--SASHSTTYNNLRINNDGIVELLHKHFDALPASSAKRILGEMMNDPALKDI 1704
Db 2149 TVTQLAAFTKSMSSLODDRVR--IDEAKKWERKFSDAIQSKEEBIRLKE--DNCSVLKQD 2205
QY 1705 IKLOSTPFSASVSM--ELKDGL--REOTEKAILDGKVGRE-----1742
Db 2206 LRQW-----SIHMEELKINISRLBHDQIWESKAQTEVLOQKQVCDTLQGENKELLS 2257
QY 1743 -----EVGLFODRNLNLRVKSVSQSVSKSEBGFNTPALILGTSNSAAMS 1788
Db 2258 QLETRHLYHSSQNELAKLSLKDQLDLSNLSLEKCKBQK-----GNLEGIIRQ 2311
QY 1789 ERNIGTINFKYQO---DQNTPRRPT--LEGGIAQANFQVASALTDLKKEGLEM 1836
Db 2312 EADIQNSKFSYEQLTDLQASRELTSRLHBEINNKQIKIISLLSG--KEEAQV 2363
RESULT 11
AC0304
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AC0304
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Ruthersford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0304
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2535 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175
C:Genetics:
A:Gene: YPO2490
```

```
Query Match 2.6%; Score 241; DB 2; Length 2535;
Best Local Similarity 19.4%; Pred No. 0.0022;
Matches 386; Conservative 232; Mismatches 659; Indels 710; Gaps 94;
QY 265 ISARFPKPLTAVAESVLEGTDTTQSPKPSQMLKSGAGVTPLATVLDKGLQ-----L 318
Db 416 LSVEASGKITAGKNI---SGTAQFTAHRLDLSGQTSQSDITLIAQGGEDLTGAELL 472
QY 319 APDNPALNTLKQTLGKDTQHYLAHASSQSGHLLLDNKGHLFDIKSTATSVLHNS 378
Db 473 ASDRLSAATTALLRT---DNASLIAEQITLDAQ---ALSNVGGI---IAHTGTTFDNL--N 522
```

```
QY 379 HPGEIK---GKLAQAGTGSVSD--GKSGKISLGSQTQSHNKTMLSQPGEAHRSLLTGIW 433
Db 523 LPGDVNRGTTLLSSGTLSQLAESLNSNGSLGAGVQSDGR--LTEIGDLRVTTRODLI 580
QY 434 QH-----PAGAARQGESIRLHD-----DKIHI-----456
Db 581 AHGQTLAAGAMALTGSRIDLADSYTOAREMTLTANRGDIDISTQBATVLAJLDTLSINTAQL 640
QY 457 -----LHPELGWV-----QSADKDTHSQLSRQADGKLYALKNRTIQLN 495
Db 641 NNQGGTLAGNTLALDQDFNQSGQVTSQDLTIDLRQDFSHQAGSTLOAGR--LTLTSL 699
QY 496 SDNKSSEKLVDKIKSVSVQORQGVAILTDPGRHKMSIMPDLSDASPESHISLSHFADAH 555
Db 700 G-----AVTNDGQL--VAGGTLSTHSDSLNS-----724
QY 556 QGLHKGSELEAQSVASHGRLV-----ADSE-----GKLFSAALPKQDGNELKMKAM 605
Db 725 GNLIATQAEALNATGALINHGEIITLGGLOTDTNLTFTNGSIIISA-----EATLNAR 775
QY 606 PQHALDEHFHGHQISGFFHDDHQLNALVKNFRQOHACPLGNNDHQHFPWNLTDLVI 665
Db 776 ERIT---NSGPDALIGA--TDENGLTALL-----APVI 803
QY 666 DNQGLHHTNPEP--HEILDM-----GHLGSLA-----LQEGK--LHVPDQ 702
Db 804 ENSDTVHTTDTAPTITLGMGTVILAGGHARDGHYASAAQVNLNLSGIESKMDLIYATT 863
QY 703 LTRG-----WTGAESD-----CKQLKGLDGAAYLLKDGVEKR 735
Db 864 LTRSRHILTANTDPIVADTVGTATVTAENPDIPGRYAEPDPCGADNSDI-----915
QY 736 LNIQSTSIKHGTENFSLPHVRNKPDPGDLQGLNK--DDKAQAMAVIGNKYKALYT--E 793
Db 916 -----GTEYTSVI-----AYNGIDQISPEAQLLA--CGN-----LTPQ 946
QY 794 KGDIRSFOIKPGTQ--QLERPAQTLSREGISGELKDI-----HVDHKQNLVALTHE 842
Db 947 VGTLENFWSKVSQAQGEIDLTGVTLOQDQGWGDQOQLMEQTTSQSSVWRYRTYKGLWREWG 1006
QY 843 GEVPHQPREAW-----ONGAESSSWHKALPQSESKLSLSDMSHEHKPI 886
Db 1007 PEVSERATSEYASFTAKTILSGSTTINNGANPGA---TAPPADRONSGKDLAVEFNGI 1062
QY 887 A-----TEDGSOHQLKAGGWHAYAAPE-----RGP---LAVGTSGSQTVFNRL-- 927
Db 1063 SLTQPNGLLYQFTTDHTVGVGGYLIETHAPANLNWRGSDYVLQQLNNDPDIKFLG 1122
QY 928 -----MQGVKGKVPISGSLTVKLSAQTGTMGAERKVSSEKSEIRA-----YAFNPMS 978
Db 1123 NAYEQRLVRDQVL-----ALTQAVASDYRSQAQEQEALFAAGLEYSKAFNIALG 1172
QY 979 TPRPIKNAAYATQH-----GWQREGIKFLYEMQGALIKQLDAHNVHNAPODLQSK 1031
Db 1173 THLSAEQMAALTHNIVLMETRDVAGQTVLVPVVLAGVKPGDLOANGALIAENISL---- 1229
QY 1032 LETLIDLGEHGA--ELLNDMKFRDEQSATRSVTVLGQHQGLKSNGB-----INSEFKP 1085
Db 1230 TEVOGFTNAGAITATNDLK-----ISMAQDIT--LNNRGLLQAGGDMQLSTLNSIDL 1281
QY 1086 SPGKALVQSFNVNRSGDLSKLOQAVHATPPSAESKLQSMGLHF-----VSAGV 1135
Db 1282 TSARINATNLQD--SGRDVILRTDQAQLSSDNGAVSRDQITLGLPLASINVSNNAINTGR 1340
QY 1136 DMSHQKEIPLGRORDPNDKTLTKSRLILDTVTIGELHELAKKLVDHDKPDADQIKQ 1195
Db 1341 DFIMQASLNVGQ-----DLQVTTGGDWQLETV-----Q 1369
QY 1196 LRQCFDTLREKRYESNPVKHYTDMGFTHN--KALEANYD---AVKAFINA-----1240
Db 1370 TRDQIST--HDGRGSATSEHRLHGVSENVGALTANVDNLTAVCANINATLEVQAQNI 1427
QY 1241 -----FKKEHHGVNLTTRTVLESQ---GSAELAKKLKNTL-----1272
```

Db	1428	SLSNATDLSLHVTGESSKRHTSSVNLVDETLGSQLNATGDNLOAAQDITLRASAVOTD	1487
Qy	1273	--LSLDSG--	1296
Db	1488	GALTAAAGGDVLLTTQTEQHDQRNHTGLSKGIASSLTWRTEDSLSQTLAVGSMLSAGSI	1547
Qy	1297	PTLSKKVPVPVPGAGITLDRAYNL--	1331
Db	1548	DVSGKNI--AVMGSNVVAQDISLRAQENITVGTAAQSSSESHLPQKKSGMLMSTGIG	1604
Qy	1332	VSFERD--	1362
Db	1605	VTVGSSSTKMTDGGQSISVSGSTVGSVLGNVMTAGEDLTVQGAELVLAGKDINLTGKNVS	1664
Qy	1363	AGNASDWLSAKHKI-SPDLRIGAAVSCITLQCTQLNSL-KFKLTDEFLPGFIHGLTHGTLT	1420
Db	1665	LLAENGLTQSHVVEQKQSGELTALSAGVSAVNTVTTAKAASESSGRLGAL--	1718
Qy	1421	PAELLQKIBHQMKQGSKLTPFSDVTSANLDRAGINLNEDGS--KPNGVTVARVSAGLSAS	1478
Db	1719	-----QGVKAALN-----GVQAVQAGQLVQAEFGDAASMFGLISAS	1753
Qy	1479	ANLAAGSRERSTSGQ-----FGSTTSASNNRPTFLNGVQAGAN-----	1517
Db	1754	-----LGSQKSSSEHQHQTHVGTGSLTAGNNLTINATGEGNAANSNGDIVVQSGLOAQGD	1809
Qy	1518	--LTAA-----LGVAH-----SSTHGBKVPGIFPAFTSTNVSAALALDNRTSQSISLEL	1564
Db	1810	TTLDARADVLLGAANTQKTDGNSSGSGGVV-----SLGISGA-----SSGLSIFANA	1859
Qy	1565	KRAEPTVNSDISELSTLGLGHFKDQSAATKMLAALKELDDAKPAQLHLQOHFSAKD--V	1622
Db	1860	NKQGSSEHGDTSWTEFT-----LDSGGLSL-----YSGRDTSL	1894
Qy	1623	VGDE-----RYEAVRNKLKLVIRQQAADSHSMELGSASHSTTYNNLS--RINNPGIVE	1673
Db	1895	VGAQVSGETVKVEVRDL--LLSQQSDSDNYDAKQNSVVGGSFSGMGTGSIINGSQD	1952
Qy	1674	LLHKHFDALPASSAKELGEMWNPALKDIIKQLQSTPFPSSASVSMELKDGUREQTEKA	1733
Db	1953	KLHSNFDSV-----QEQTGHIFAGSGGFDITVGGHTQLDGA	1987
Qy	1734	ILDGKVGRE-----EVGVL-FODRNN--LRVKSVSVSQSV-----SKSEGFNTPALLIGT	1780
Db	1988	VIGSTATDNKTLDTGLGLGSDIDNQADFVKEHQSGVIGTGGNIGSQFQVGNMANGLLVGA	2047
Qy	1781	SNS-----AAMSBERNITGTFNK--YQGDQNTPRRFTLEGGIAQANPQVASALTDLK	1830
Db	2048	NNEGHADSTTHAAVSE---GTITVRDQDQOQNVD---DLSDRVEQAN-NALSPFIDXE	2099
Qy	1831	KEGLEMK 1837	
Db	2100	KEQNRK 2106	

RESULT 12

T13564
Microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N:Alternate names: hypothetical protein EG:4924.1
C:Species: *Drosophila melanogaster*
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiamnakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A:Reference number: Z17689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:07:11 ; Search time 26.0314 seconds
(without alignments)
3676.525 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAAHNNPQVASALTDLKKEGLEWKS 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244.5	2.6	3259	1 GOB1_HUMAN	Q14789 homo sapien
2	240.5	2.5	5560	1 SPEN_DROME	Q88x83 drosophila
3	232	2.5	1577	1 HLYA_PROMI	P16466 proteus mir
4	229.5	2.4	2541	1 TLN1_HUMAN	Q9V490 homo sapien
5	221	2.3	1608	1 HLYA_SERMA	P15320 serratia ma
6	219.5	2.3	2541	1 TLN1_MOUSE	P26039 mus musculu
7	218.5	2.3	5262	1 MLT2_HUMAN	Q14686 homo sapien
8	217.5	2.3	1957	1 SPOF_SCHPO	Q10411 schizosacch
9	214	2.3	1781	1 AK12_HUMAN	Q02952 homo sapien
10	205	2.2	2660	1 YEBJ_ECO57	Q8X8V7 escherichia
11	204	2.2	1447	1 GOA3_MOUSE	P55937 mus musculu
12	204	2.2	2845	1 APC_MOUSE	Q61315 mus musculu
13	202	2.1	5171	1 BPEA_HUMAN	Q94833 homo sapien
14	199.5	2.1	3590	1 FHAB_BORPE	P12255 bordetella
15	199	2.1	1191	1 CING_MOUSE	P59242 mus musculu
16	199	2.1	2319	1 AKA6_HUMAN	Q13023 homo sapien
17	198.5	2.1	835	1 INVO_PONPV	P14708 pongo pygma
18	197.5	2.1	1979	1 TPIA_HUMAN	Q15643 homo sapien
19	196.5	2.1	1902	1 P2P_LACPA	Q02470 lactobacill
20	196.5	2.1	5038	1 PCLO_MOUSE	Q9GYX7 mus musculu
21	195.5	2.1	4377	1 ANK3_HUMAN	Q12955 homo sapien
22	195	2.1	3321	1 PCN2_HUMAN	O95613 homo sapien
23	195	2.1	5938	1 MAC4_HUMAN	Q956k2 homo sapien
24	194	2.1	1861	1 MAP2_RAT	P15146 rattus norv
25	194	2.1	3358	1 PGCV_MOUSE	Q62059 mus musculu
26	194	2.1	5430	1 MACF_HUMAN	Q9UPN3 homo sapien
27	193.5	2.0	1902	1 P3P_LACLC	P15292 lactococcus
28	193	2.0	2238	1 GOA4_MOUSE	Q91VW5 mus musculu
29	191.5	2.0	2738	1 PGCV_RAT	Q9ARB4 rattus norv
30	190.5	2.0	8545	1 ANCL_CABEL	Q9N4M4 caenorhabdi
31	190	2.0	2442	1 CEP2_HUMAN	Q9BV73 homo sapien
32	189	2.0	1902	1 P2P_LACLC	P15293 lactococcus
33	189	2.0	2175	1 HMCU_DROME	P10180 drosophila

RESULT 1

ID	GOB1_HUMAN	STANDARD;	PRT;	3259 AA.
AC	Q14789; Q14398;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)			
DE	(Golgi complex-associated protein, 372-kDa) (GCP372).			
GN	GOLGB1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
EX	MEDLINE=94187728; PubMed=7511208;			
RA	Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,			
RA	Renz M.;			
RT	"Molecular genetic analyses of a 376-kilodalton Golgi complex membrane			
RT	protein (giantin).";			
RL	Mol. Cell. Biol. 14:2564-2576(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
EX	MEDLINE=94257116; PubMed=8198703;			
RA	Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,			
RA	Renz M.;			
RT	"Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as			
RT	target of antibodies in patients with rheumatic diseases and HIV			
RT	infections.";			
RL	J. Autoimmun. 7:67-91(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
EX	MEDLINE=95100974; PubMed=7802676;			
RA	Schda M., Miumi Y., Fujiwara T., Nishioka M., Ikehara Y.;			
RT	"Molecular cloning and sequence analysis of a human 372-kDa protein			
RT	localized in the Golgi complex.";			
RL	Biochem. Biophys. Res. Commun. 205:1399-1408(1994).			
CC	-!- FUNCTION: May participate in forming intercisternal cross-bridges			
CC	of the Golgi complex.			
CC	-!- SUBUNIT: Homodimer; disulfide-linked.			
CC	-!- SUBCELLULAR LOCATION: Golgi; membrane-associated.			
CC	-!- DISEASE: Antigen in chronic rheumatoid arthritis and in the			
CC	autoimmune disease Sjogren's syndrome.			
CC	-!- SIMILARITY: Belongs to the golgin family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X75304; CAA53052.1; -.			
DR	EMBL; D25542; BAA05025.1; -.			

ALIGNMENTS

DR PTR: A56539; A56539.
 DR PIR: 152300; 152300.
 DR Genew; HGNC:4429; GOLGB1.
 DR MIM; 602500; -.
 DR GO; GO:000139; C:Golgi membrane; TAS.
 DR GO; GO:0005795; C:Golgi stack; TAS.
 DR GO; GO:0016021; C:Integral to membrane and biogenesis; TAS.
 DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 3236 3256 POTENTIAL.
 FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
 FT DOMAIN 48 593 COILED COIL (POTENTIAL).
 FT DOMAIN 677 1028 COILED COIL (POTENTIAL).
 FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).
 FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
 FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).
 FT DOMAIN 2420 2423 POLY-GLU.
 FT DOMAIN 2993 2996 POLY-SER.
 FT CONFLICT 1 39 MISSING (IN REF. 3).
 FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
 FT CONFLICT 1765 1765 D -> G (IN REF. 3).
 FT CONFLICT 2950 2950 H -> D (IN REF. 3).
 FT SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;
 Query Match 2.6%; Score 244.5; DB 1; Length 3259;
 Best Local Similarity 18.6%; Pred. No. 0.0024;
 Matches 389; Conservative 311; Mismatches 816; Indels 577; Gaps 92;
 QY 23 GHGVNLOQSSSSPQNAASLAEGKRGKMPRIHQ--PSTAADGISAAHQKKSFLR 80
 DB 603 GEGIA-----PIKMKVLEDTGQDFPLMNEESSLPFAVEKEQASTHEQRTSEI- 652
 QY 81 GCLGKTFSSAPQGPQGTTHSKGATLRDLLARDGGETOHEAAPPDAAR----- 129
 DB 653 -----SLNDAGVELKS--TKQDGD-KSLSAVDPDLCQCHQDELERLKS 691
 QY 130 -----LTRSGVKRRNDDWAGRMVVGSGGDEKVPTQQRKHOLN-----FGQMR 175
 DB 692 QILELEINFHQAQIEYKNDLDEKA-----KEISNLNOLIEFFKKNADNNSAFTALSEER 746
 QY 176 QTMLSKM-----AHPASANAGDRLOHSPPHIPGSHHEIKPEPVGTSK 218
 DB 747 DQLLSQVKELSMVTELRQAQVKLEMLNLAERQRRLDYES---QTAHDNLLTTOIHLSI 803
 QY 219 ATTAHADRVETIAQED-DDSEFQQLHQORLARENPFPQPKLVY-----ATPISARFP 271
 DB 804 EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRSLQSQLOKNESEVLEGAERVRISSKVEE 863
 QY 272 KLTAVASVLEGTTDTTQSPKLP-----QSMKLGSGAGVTPPLAVTLDKGLQLAPDN-- 322
 DB 864 LSOALSQKELEITKMDQLLEKKRDVETLQOTIEKDDQVTEISFSMTKRWQVNBKFS 923
 QY 323 -----PPALNTLLKOTGLKDTQHYLAHASSDGSOHLLDNKGLHFDIKSTATSYV 374
 DB 924 LGVEIKTLKQNLNLSRABEAKKEQVEEDNEVSGLKQN-----YDEMSPAGQISK 974
 QY 375 LHNHPGEEI-----KGKLAQAGTGSVSDGSKGIS---LGSQTSHNKTMLSQP-- 421
 DB 975 EELQHEFDLLKKEQKRKLQALINRKELLQVRSLBEELANLKDSEKKEIPLSETER 1034
 QY 422 GEHRSLITGIWHPAGAARPOGESIKLHDDKIHILHPELGVWQSDAKDTHSOLSRQADG 481
 DB 1035 GEVEEDKENKEYSEKVTSCQETIYL---KQTISEKEVEL-QHTRKDLKEELAAEEQF 1090
 QY 482 KLYALDKNRTLQN-----LSDNKSSEKLVDKIKSYVD--QRGOVAILTDTTGR 528
 DB 1091 QALVKQNNQLQKTNQIDLLQAEISENQ--IIQKLITSNTDASGDGVALVKET--- 1144
 QY 529 HKMSIMPSLDASPESHISLSLHFADAHQGLHGKSELEAQSVAISHGLRVVADSEGLKFS 588
 DB 1145 --VVISPPCTGSSE-----HWKPELEEKILALEKEK---EQLOKKLQE 1182

QY 589 A-----AIPKQGD-----GNELKMKAMPQHALDEHFGHDHQISGFFPHDHGQNALVKN 637
 DB 1183 ALTSRAILKKAQEKERHLREELKQKDDYNRLQEQEQDEQSKENENIGDQLRLQIQVRE 1242
 QY 638 NF-----POHACPLGNDHQHPGWNLTALVIDNQLGLHHTNPPEHLLDMGLGSL 690
 DB 1243 SIDGKUPSTQQESC-----SSTFG--LEEPFKATE--QHTQTP-----VLE----- 1281
 QY 691 ALQEGKLYHFDQLTGKMTGAESDCKOLKGLDGAAY--LLKDGVEVKR-----LNINQSTSS 744
 DB 1282 -----SNLCPDWPSESDASALOGGTVAQIKQAQLKEIAEKVELELKVSTTSE 1331
 QY 745 IKHTENVFSLPHVRNKPPEGDALQGLKDDKAQAMAVI GVNKYALTEKGDIRSFQIKP 804
 DB 1332 LTKKSEVFLQEQINK-----QGL-----EIES--LKT 1358
 QY 805 GTQOLRPPAOTLSREGISGELKDIHVDHKONLYALTHEGEVFHPRE--AWQNGAESS 861
 DB 1359 VSHAEEVHAESLQOKLESSQLTAGLEHLRELQPKDELKLSKKEEDVSYLSGQUSEK 1418
 QY 862 WHKLA-----LPQSESKLSLDSHEHKPIATFEDSGHQKAGGWHAYAAPERGLAVG 916
 DB 1419 EAALTQIQTETIEEDDLIKALHTQLEWQ--AKEHDERIKOLQVELCEMKQKPEB---IG 1472
 QY 917 TSGSQTVFNLMOGVKVIPIGSLTVKLSAQGTGMTGAERKVSSEKSERI---RAYAF 973
 DB 1473 EE-----SRAQKQIQRL-----QAALISRKALKENKSLQEELSLARGTIE 1514
 QY 974 NPTMTSTPRIPKNAAYATOHQWQRE-----GLKPLYEMOGALIKOLDAHNVHNAFPD 1027
 DB 1515 RLTKSL-----ADVESQVAQNKEDTVLGRALIQEERDKLITEMD-----RELLENQS 1564
 QY 1028 LQSKLETLDLGEHGAELLNMDKFRDELEQSAIRSVTVLGQHQGVLSKNGEINSEKFPSP 1087
 DB 1565 LSSSCESLKLALSG--LTDKELKVKIESLSSKIAESTEWO--EKHKELQKEY----- 1615
 QY 1088 GKALVGSFVNVRSGDLSKSLQQAIVATPPSAESKLQSMLGHFVSAGVDSH-----OKG 1142
 DB 1616 -ELLOSQ-----ENVSNEAERIQHVVEAVROEK-QELYGKLSTEAANKETEKQJQEA 1667
 QY 1143 EIPLGRQRPNDKALTALSKRLILDTV-----TIGELHELADKAK-----LVSDHKPDADQ 1192
 DB 1668 EQMEEMEKMKRPAKSKQKQKILLEENDRLRAEVHPAGDTAKECMTLLSSNASMKEE 1727
 QY 1193 IKQLRQOQDPLREK-----RYESNPVKHYTDMGTTHNKALAEANYDAVKAFINAF 1241
 DB 1728 LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE--DNVSKQANLEAT----- 1776
 QY 1242 KXHHGYNLTTRTVLSQGSAAELAKKLKNTLLSLDS-----GESMSFSRSYGGGVSTVP-- 1295
 DB 1777 EKHDNQTNTVEGTQSIPIGTE-----EQDSLMSMSTRPTCSSESVPSAKSANPAVSKDFSS 1831
 QY 1296 -----VPTLSKKVPVPVPIPGAGITLDRAYNLSFSRTSGG-----LNVSGRGGV 1340
 DB 1832 HDEINNVLOQIDQKERI-----AGLEBEKQKNEFQSTLENEKNTLLSQISTKDGEL 1884
 QY 1341 -----SGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPDLRIGAAVSGTLOG 1392
 DB 1885 KMLQEBVTQNNLNLQOIQEELSRTVKETAESEKD-----DLEE 1924
 QY 1393 TLQNSLKFKLTEDDELPGFHLGHTGLTTPAELLQKIEHOMKQGSKLTFSDV----- 1444
 DB 1925 RLMNQLA-----ELNGSI--GNYCQDVDTAQIKNELLESEMKNLAKVCSEBEKQOLVK 1377
 QY 1445 --TSANLDLR-----AGINLNEDGSKPNV-----TARVSAGLS 1476
 DB 1978 EKTQVSESEIRKEVLEKIQGAQ--KEPGNKSHAKELQELLKEQOEVLQKQDCIRYQEKIS 2036
 QY 1477 ASANLAAGSRERTTSQGFSTTSASNNRPTPLNGVGAGANLTAALGVASHSTHEGKPVG 1536
 DB 2037 A-----LERTVKALEFVQTSQKDLKTEIKEN-----LAQAVEHRKKAQELAS 2079
 QY 1537 IFFAFTSTNVSAALAL-DN-----RTSQSISLSELKRAEPTVTSNDIS-ELASTLGHF 1586


```

Db      825  ---GSPFRFRYESSASSLSQSRSSASSFRHQNNSNDCCSPINTPGASSGSISSASNLIN 881
Qy      778  QAMAV-----GVNKYLALTEKGDILRSQIKPGTQOOLERPACTLSREGISGELKDIHVDHK 833
Db      882  QSTISINIGTNCASAM-----PAPSLASAVVS-----CNVNAS 916
Qy      834  ONLYALTHEGEVFPQPREAWONGAESWHKIALPOSESKLKSLDMGHEHKPIATFEDGS 893
Db      917  GTVPAST-----SMPGVSSSS-----SSLPMSPAL-----AQRHVMVENARQTV 957
Qy      894  QHQLKAGWHAYAAPERGLAVGTSGQTVNRL-----MOG-----VKGK 934
Db      958  DCDNEVGRILFRSSEE---VSGGAGNSTQPEDVRCDSPVTARQGSVAVNCFGTGTAAVGE 1014
Qy      935  VIPSGSLTVKLSAQTGCTGAERKVKSKTSERIRAVAFNPTMTSTPRPKNAAVATQHW 994
Db      1015  SIDGTILNNQITGAEGFTSGGILRRRCGK-----TPK----- 1050
Qy      995  QGREGLKPLYEMOGALIKQIDAHVNRHNAPODLQSKLETLDLGEHGAELNDMKRFRDE 1054
Db      1051  ----DLHPVHNRQLAEQVE-----ECFSSDEG---VVSPPKRIKMD 1087
Qy      1055  LEQSATRSVTVLGHQGVKNSNGEINSEFKSPKALVQSPNV-----NRSGQLSKSLQ 1109
Db      1088  YHHHHHS-----NAGSVSTGEHSSINKDSP--LLLSNCDVIHDPILNR-----KSEI 1133
Qy      1110  QAVHATPPSAESKLQSLMGLHFVSA--GVDMHQKGEIPLGRQRPNDKTLTKRLIILD 1167
Db      1134  RRVSETSGSPS--IKFPHLPSPAPQSLMLSCRRPSIDVGA-----LSALSSSAP--- 1182
Qy      1168  VTIGELHELADKALVSDHKPDADQIKQLRQOFDTLREKRYESNPVKHYDTMGTHNKAL 1227
Db      1183  ----RHGIVGASSMDQHHMWNASAAKRRVTTTMOQPSSSS-----ITNSSS 1226
Qy      1228  EANYDAVAFINAFKKEHGVNLTRTVLESGGAELAKLKNLLSLDSGESMSFSRSY 1287
Db      1227  GSGLLGSSSLTPA-DEYHHVSVSRGRGHQHLHSHSHEAS-----GGESADGSRP- 1273
Qy      1288  GGGYSTV-----FVPTLSKKYVVP-----VIPGAGITLDRAYNLSPSRTSG 1328
Db      1274  --GTFCLDERPEVLTFEPRLPPPRVRERTRDVMMLPLPKFGVLF---FQQQSSRSSG 1328
Qy      1329  GLNVSPGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPLDRIGAAVSG 1388
Db      1329  G-----GG-----AGSVYLOQLGGSGTGGLGC-----IGAASSS 1358
Qy      1389  TLQGTQLNSLKFKLTEDLPFGIHLGHTLTPAELLQKIEHQMKQSKLTFV-----D 1444
Db      1359  AC--SLNNS--SLNASQGMGSCSGSTPLSPSSRYWRSSHQNNHQOQNSQQLHGS 1413
Qy      1445  TSANLIDRA-----GINLNEGSKPENGVA-----RVISAG----- 1474
Db      1414  SSSNTCLMASPARPRSUSNDSDDVPCQNAAGGSPSLDERLRNFENYERWGSSSREHI 1473
Qy      1475  -----LSASANLAAG--SRERSTTSQFGS-----TTSASNNRPTFLN----- 1510
Db      1474  SGHTFPSSATPQWLSMHNLTGLNSHQTSSAGNSNSSSGTVSSASNSRHKFLDIDEL 1533
Qy      1511  -----GVGAGANL-TAALGVVAHSSTHEGK 1533
Db      1534  QPSDIVKSVLAKSVFDDDFQRLNKNQWYDFSSSDFALGSSSNIVTGSSLVANVSRHPGG 1593
Qy      1534  PVGIFPATSTINVSAAALDNRQTSQISLELKRAEPTVSDISELTSTLGHFKFDASATK 1593
Db      1594  P-----CSGNTSPALP-----NLAATKATPIIICGSGGLNGTGS--KSAGLLQ 1635
Qy      1594  MAAELKELDDAKPAQOLHILOQHFSAKDVGVDERVEAVENLKKLVIRQQAADSHSMELGS 1653
Db      1636  RLSLSLPMN--SPQASMGFPYNSPSPSVGG-----VTACLGQLTKPAAPGTASAGL-- 1685
Qy      1654  ASHSTYNNLRINNDGVELLHKHFDALP-ASSAKRLGEMMNDPALKDIIKOLQSTP 1712

```

```

Db      1686  -SGGTAASSSPAAANSQPTKGLQYPPFSPHPPLPNTAAPPVAPQAPPPLPMGKQSRUTG 1744
Qy      1713  FSSA---SVMELKDG-----LREQTEK-AILDGKVGREEVGLFQDRNNLRVK-SVSVS 1762
Db      1745  QSSGNLTKLSVDPDGPQSSPARVOLQKSASVPGSTNVGAPSSLSLDDSTTASVTSASIS 1804
Qy      1763  QSVSKSGEGNTPAILL-----LGTNS-AAMSMERNIGTINFKYGDQ 1803
Db      1805  SSTSGNSSLTSAAIHVQPKQSTFVEEHTKKSGTSTQSSSSSSSKKISSTHDKLHSHK 1864
Qy      1804  N 1804
Db      1865  N 1865

RESULT 3
HLVA PROMI
ID HLVA PROMI STANDARD; PRT; 1577 AA.
AC P16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN HPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC STRAIN=Isolate 477-12;
RX MEDLINE=90170827; PubMed=2407716;
RA Uphoff T.S., Welch R.A.;
RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent
RT hemolysin genes (hpmA and hpmB) reveals sequence similarity with the
RT Serratia marcescens hemolysin genes (shlA and shlB).";
RL J. Bacteriol. 172:1206-1216(1990).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
CC REQUIRES HPMB FUNCTION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA
CC MAY BE RESPONSIBLE FOR PORE FORMATION.
CC -!- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

-----
CC EMBL; M30186; AAA25657.1; --
DR PIR; A35140; A35140.
DR InterPro; IPR008638; Haemagg_act.
DR Pfam; PF05860; Haemagg_act; 1.
KW Hemolysis; Toxin; Outer membrane; Signal.
FT SIGNAL 1 29
FT CHAIN 30 1577 HEMOLYSIN.
SQ SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

Query Match 2.5%; Score 232; DB 1; Length 1577;
Best Local Similarity 18.9%; Pred. No. 0.003;
Matches 329; Conservative 238; Mismatches 614; Indels 564; Gaps 84;

Qy 261 VATISARFQKLTAVAESVLEGTDTTQ-----SPLKPSQMLKSGAGVTPPLAVTLQK 314
Db 197 IAPRIDGR--GKITAABISAFTQNTFSQHFIDILUSSQKPVSAIDSYFFG-----SMQSG 248

```


CC similarity).
 CC -|- SUBUNIT: Binds with high affinity to vinculin and with low
 CC affinity to integrins (By similarity).
 CC -|- SIMILARITY: Contains 1 PIRM domain.
 CC -|- SIMILARITY: Contains 1 I/LWEQ domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch.
 CC -----
 CC EMBL; AF078828; AAD13152.1; -
 CC EMBL; AF177198; AAF23322.1; -
 CC EMBL; AF178534; AAF27330.1; -
 CC EMBL; AF178081; AAF27330.1; JOINED.
 CC Genew; HGNC:11845; TLN1.
 CC MIM; 186745; -
 CC GO; GO:0005925; C:focal adhesion; NAS.
 CC GO; GO:0005200; P:structural constituent of cytoskeleton; NAS.
 CC GO; GO:0006928; P:cell motility; NAS.
 CC GO; GO:0007016; P:cytoskeletal anchoring; NAS.
 CC InterPro; IPR00299; Band 4.1.
 CC InterPro; IPR002558; ILWEQ.
 CC Pfam; PF00373; Band 41; 1.
 CC Pfam; PF01608; ILWEQ; 1.
 CC ProDom; PD011820; ILWEQ; 1.
 CC SMART; SM00295; B41; 1.
 CC SMART; SM00307; ILWEQ; 1.
 CC PROSITE; PS00660; FERM 1; 1.
 CC PROSITE; PS00661; FERM 2; 1.
 CC PROSITE; PS00657; FERM 3; 1.
 CC PROSITE; PS00945; ILWEQ; 1.
 CC Structural protein; Cytoskeleton.
 CC DOMAIN 86 403 FERM.
 CC FT DOMAIN 2340 2533 I/LWEQ.
 CC FT CONFLICT 824 824 G -> R (IN REF. 2; AAF27330).
 CC FT CONFLICT 1227 1227 S -> L (IN REF. 1).
 CC FT CONFLICT 1549 1549 P -> A (IN REF. 2; AAF27330).
 CC FT CONFLICT 1604 1604 Q -> K (IN REF. 2; AAF27330).
 CC FT CONFLICT 1701 1701 E -> Q (IN REF. 2; AAF27330).
 CC FT CONFLICT 1718 1718 H -> N (IN REF. 2; AAF27330).
 CC FT CONFLICT 1966 1966 A -> R (IN REF. 1).
 CC FT CONFLICT 2256 2256 MISSING (IN REF. 2; AAF27330).
 CC SEQUENCE 2541 AA; 269717 MW; 47FD9B22BCF47296 CRC64;
 Query Match 2.4%; Score 229.5; DB 1; Length 2541;
 Best Local Similarity 19.5%; Pred. No. 0.0083;
 Matches 449; Conservative 297; Mismatches 846; Indels 707; Gaps 114;
 QY 11 KAAVHTAHNPVG---HG-VALQ--QSSSSSPON-----AAASLAEGKRGKMPR 56
 DB 428 KSTVLOQYNRVGVHGSVALPAIRMSGASGPFVQSGMPPAQQITSGMERGHMP 487
 QY 57 IHQPSTADG-----ISAHQOKKSFSLRGCLG---TKFSRAPOQPGTTHSK- 103
 DB 488 LTSQAQALTGTINSMQAVQAAQATLDDFDLPGLQDAASKANRWKKNKDESKHEHSQV 547
 QY 104 -----GATLDLLARDGQTOHE-----AAPDAARLTR--SGGVK--RRNDDM--AGRP 148
 DB 548 DAITAGTASVNLNAGDPAEDTYAVGCAVTTISNLTMSRGVKLLAALLEDEGGSGRP 607
 QY 149 MVKSGGDEKVPYTOQRHQLNFGMRQTMLSKAHFASANAGRLQH-----SPPHIPG 203
 DB 608 LLQAAKGLAGAVSELLRSQAQAPASAPRQNLQAAGNVGQA--SGELLQIGESDTPHFQD 666
 QY 204 SHHEKEEVPVSTKATTAHADRVETIAQEDDDSEFQHLHQRLARERENPPPKLGWAT 263
 DB 667 ALMQI--AKAVASAAALVLKAK--SVAQRTSDSLQ-----TQVIAATQCALSTSLQVACT 720
 QY 264 PISARFQPKLTA-----VAESVLEGTDTTQSPKQSMKLGSGAGVTPLAVT 310

DB 721 KWVA---PTISSPVQEQQLVEAGRLVAKAVEGCVSASQAATEDQGLLRGVGAAT--AVT 775
 QY 311 LDKGKLQALPDNPALNTLLKQTLGDKTOHYLAHASSDGSQHLDDNKGHLFDIKSTAT 370
 DB 776 -----QALNELL-----QHVKAHATGAGPAGR-----YD-QATDT 804
 QY 371 SYSVLHN-----SHPGEIKG---LAQAGT---GSVSDGKSGKISLGSGTQSHNKTMLS 419
 DB 805 ILTVTENIFSMGMDAGEMVGOARILQAATSDLVNAIKADAE-----GESDLENSRKLSS 858
 QY 420 QP---GEARSLTLGTIWOHPAGAAARPOGESIRLHDDKIHILHBPGLGWQADKDTHSQLS 476
 DB 859 AAKILADATATKVV---EAAKGA-----AHPD-----SEBQQQRLR 891
 QY 477 ROADGKLYALKDNRTLONLSDNKSSEKLVDKIKSYSDQRQVAILTDTTPGRHKMSTPS 536
 DB 892 EAHEG---LRMATNAAQ-----NAKKKLQVRLK- HAAKQAAASATQITIAAQAHAAS-TPK 943
 QY 537 LDASPESHISLSLH- FADAHQGLLHGKSELEAAQVAISHGRLVVVADSE-----GKLFS 589
 DB 944 ASAGPQLLVQSCKAVAEQIPLLVQGVRSQAQPDSPSAQLALIAASQSFQPGKMWAA 1003
 QY 590 A---IPKQGDGNELMKAMPQHLDHDFHGDHQSISGFHDDHGGQNLALVKH----- 637
 DB 1004 AKASVPTIQD-----QASAM-----QLSOCAKNLGTALABLR 1035
 QY 638 --NFRQOHAC-PLGNDHFHPGNL-----TDALVIDNQLGLHHTNPPEHIL-----D 683
 DB 1036 TAAQQAQACGPLEMSALSVMQNLKDLQEVKAAARDGKL-----KPLPGETWEKCTQD 1090
 QY 684 MGH-----LGSALQEGKLHYFDQLTKGWTGAESDCQKQKGLDGAAYLLKDG 731
 DB 1091 LGNSTKAVSSAIAQLLGEVA--QGNENYAGIAARDVAGG---LRSQAQAARGVAALTSDP 1145
 QY 732 EVKELNINQSTSSIKHGTENVFSLPHVRNKPDPGDALQGLNKDKQAQAMAVIGNKYL-- 789
 DB 1146 AVQAIVLDTASDVLDKASSLIEEAKKAAGHPDPEEQORLAQVAKAVTQA---LNRVCSC 1202
 QY 790 -----ALTEKGD---LRSFQIKPGTQQLERPACTLSREGISG-----ELKDIHV 830
 DB 1203 LPQORDVDNALRAVDGASKRLSSLPSTGTFTPOE-AQSRLENAAGAGLNQAATELVQASR 1261
 QY 831 DHKNL-YALTHEGEVFPQPREAMQNGAESSSWHKLALPOSE-----SKLKSIDMSHEH 883
 DB 1262 GTPQDLARASGRFGQDFSTFLEA--GVEMAG---QAPSQEDRAQVVVNLKLGISMSSSK 1314
 QY 884 -----KPIATFEDGSQHLKAGWHYAAAPERGLAVGTSGSQTV----- 923
 DB 1315 LLLAAKALST--DPAAPNLKS---QLAAAAA---AVTDSINQLITMCTQQAQPGKECDN 1365
 QY 924 -----FNRLMQGVKGVIPGSGLT-VKLSAQTG---GMT 953
 DB 1366 ALRELETVRELLENVPVINDMSYFGCLSDVMENSKVLGAMTGISQAKNGNLPFEGDA 1425
 QY 954 GAEGKYSKFSERIRAYAFNPTMTSPRIKNAAYATQHGWQREGKGLPLYEMOGALIKQ 1013
 DB 1426 ISTASKALCGFTEAAQAAYLVGVSDP-----NSQAQOQGLVEPTQPARA--- 1470
 QY 1014 LDANVVRHNAPQDLOSKLETLDLGEHG---AELLNDMKRPRDELESATRSVTVLGQH 1070
 DB 1471 -----NQAIQMACQS-----LGEFGCTQAQVLS-----AATIVAKHT 1502
 QY 1071 GVILKSNGEINS--EFKPSPGKALVQSF-NVNRSQDLKSILQ-----QAVHAT 1115
 DB 1503 SALCNCRSLASARTTNTAKRQFVQSAKEVANSTANLVKTIKALDGPFTTEENRAQCRAT 1562
 QY 1116 PPSAE-----SKLQSMGLGH-----FVSAGVDMSHQKGIPLGR----- 1148
 DB 1563 APLLEAVDNLSAFASNPFFSSIPAQISPEGEAAMEPIVISAQTMLESAGGLIQTARALAV 1622
 QY 1149 -QRDPNDKALT-KSRILIDTV-----TIGELH-ELA-----DKAKILVS 1184

Db 1623 NRPDPSPSVLGHSTRTSDSIKKLITSMRDKAPQLECEETAALNSCLRLDQASLAA 1682
 QY 1185 -----DH-----KPDADQ-----IKQLRQED--T 1202
 Db 1683 VSQOLAPREGISQALHTEMMLTAQVEISHLIEPLAHAARAEASQLGHKVSQMAQYFEPLT 1742
 QY 1203 LREKRYESNPKHYTDMGFTH-----NKALEANYDAVKAFINAFKKEHGVNLTTRTVL 1256
 Db 1743 LAAGAASKTSLHPQOMALLDQTKTAEASALQLLYTAKEAGGNPKQAAH-----TQCAL 1796
 QY 1257 ES--QSSAELAKKLKNTLSLDGSEMSFSRSGVSTVFPVTLKKVPVFPVPGAGIT 1314
 Db 1797 EEAQVMTEAEDLTTL-----NEAASAGVGVGMVDSI--TQAINQLDRGPMGEPSGF 1850
 QY 1315 LDRAYNLSPRTSGGLNVSGFRDGVSGNIMVATGHDVMPYMTG----- 1358
 Db 1851 VD--YQTTWRTAKAIVTQVE-----MTKSNTPSEELGPLANQUTSYGLASE 1899
 QY 1359 ---KKTASAGNSDWLSAKHKISPLDRIGAAVSGTLQGTQLNSLKFKLTEDELPFIHGLT 1415
 Db 1900 AKPAVAEAENEETGSHIKHRVQ--ELGHGCAALVTKAGALQCSFSDAYTKKEL----- 1950
 QY 1416 HGTLTFAELLQKIEH--OMKQSKLTFSVDTSAN-----JDLRAGI-----NINED 1460
 Db 1951 ---IECARRVSEKVSHVLAALQAGNRGTGCITAAASAVSGIIADLDTTTFATAGTLNRE 2007
 QY 1461 GSKP-----NGV--TARVSA--GLSASANLAAGSRSTTSQO-----FGST 1498
 Db 2008 GTEFFADHREGILKTAQVLVEDTKVLVONAGSQEKLAAQASVSATIFLADVVKLGA 2067
 QY 1499 TSASNNRPTFLNGVAGANITAAALGVAHSTH--EGKPVGIFPAFTSTNVNSAALADNRT 1556
 Db 2068 SLGAEDPETQVVLINAKVAKALGDLISATKAAAGK--VGDDPAVWLKNSAKVMVTNVT 2126
 QY 1557 S-----QSISLEL-----KRAEPTVNDISELSTSLGKHF 1586
 Db 2127 SLKTKVADEATKGTALAEATTEHROELAVFCPEPPAKTSTPEDFIRMTKGI----- 2182
 QY 1587 KDSATKMLA-----ALKELDDAKPAEQHILQOHFSKADVVGDERVEAVRNK 1635
 Db 2183 -TWATAKAVAGNSCRQEDVIATNLSRRADIMWLPRACKAAVHPEVAPVRLRALHYGR 2241
 QY 1636 KLV-----IRQQAADSHMELGASH-----S 1657
 Db 2242 ECANGYLELLDHVLLTLQKPSPELKQTLGHSKRKAVGVTTELIAQAEAMKGTBWDPEDP 2301
 QY 1658 TTYNNLSRINNDGIVELLHKHFDAAALPASSAKRLGEMMNDPALKDIIKQLQSTPSSAS 1717
 Db 2302 TVIAENELGAAAEAAAKKLBOLKPRAPKPEADESLNFEQIILEAKSI-----AAA 2355
 QY 1718 VSMELKDGLREQTEKATLDGKVGREEVGVLFQDRNNLRKVSQSVSOKS----- 1768
 Db 2356 TSALVKAASAAQRE--LVAQKVGAI PANAL--DDGQWSQGLISAARVAAATNNLCEAN 2412
 QY 1769 ---EGFNTPALLGTSNAAAMERNIGTINFKYQDQNTPRPTLEG--GIAQAN----- 1819
 Db 2413 AAVQGHASQKLISSAKQVAASTAQLIVACKVKADQDSEAMKRLQAAGNAVKRASNILVK 2472
 QY 1820 -POVASALITLKKLEGLMK 1837
 Db 2473 AAKRAAFEEQENETVVVK 2491

RESULT 5

HLVA_SERMA

ID HLVA_SERMA STANDARD; PRT: 1608 AA.
 AC PL5320;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Hemolysin precursor.
 GN SHA.
 OS Serratia marcescens.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Serratia.
 RN NCBI_TaxID=615;
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
 RC STRAIN=SN8;
 RX MEDLINE=88257037; PubMed=3290200;
 RA Poole K., Schiebel E., Braun V.;
 RT "Molecular characterization of the hemolysin determinant of Serratia
 marcescens";
 RL J. Bacteriol. 170:3177-3188(1988).
 CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
 cell membranes and cause cell rupture by mechanisms not clearly
 defined.
 CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
 ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHA
 CC REQUIRES SHLB FUNCTION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M22618; AAA50323.1; --
 DR PIR; A28182;
 DR InterPro; IPR008638; Haemagg_act.
 DR Pfam; PF05860; Haemagg_act; 1.
 KW Hemolysis; Toxin; Outer membrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 1608 HEMOLYSIN
 SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;
 Query Match 2.3%; Score 221; DB 1; Length 1608;
 Best Local Similarity 18.6%; Pred. No. 0.01;
 Matches 372; Conservative 235; Mismatches 689; Indels 700; Gaps 89;
 QY 33 SSSSPONAAASLAAGKRNKMPRIHQPTAA-----DGISAHQKKSFLR--GC 82
 Db 23 AASAGAYAAIBAVANGANG---PGVSTAATGAQVDIVAPNGNGLSHNQYQDFNVNPGA 79
 QY 83 L--GTKKFSRSPQGPQGTTHSKG---ATLRLDILLARDG-----ETQHEAAA 124
 Db 80 VLNNSREAGLSQALQGLGANPLNGREASVILNEVIGRNPSSLHGQOIFGMAADYVLAN 139
 QY 125 PDAARLTRSGGVKRRNMDMAGRPVYKGGSGEDKVPYQKKRHLQNNFGQMRQTMLSKMAH 184
 Db 140 PNGISQSCGFIINTSHSLVVGNNPLVGVLOG-----YSTFGNRTLSLN--- 185
 QY 185 PASANAGDLRLQSPPHIPGSHHEIKEEPPVGSSTKATTAHADRYE--TAQEDDDSEFQOLHQ 243
 Db 186 -GTINAGGVLDLTAQKIDSRGEVIVQDFKQSNKGVTSAAINAIISGLNRVARDGTVQASQ 244
 QY 244 QRLARERE-----NPPQPKLGVAIPISARFOPKLTAV-----AESV 280
 Db 245 MPTALDSYILGSMQAGRIINIINTAQSGVYKLAGSLNAGDELKVKADIRSEVRDASSN 304
 QY 281 LEGTDTTQ-----SPLKPSMLKG-----SGAGVTPPLAVTLDKG 314
 Db 305 KNGGNDYQYRGYIYVNDRSSSTLTTRTELKGNISLVADNHAHLTATDIRGEDITLQGG 364
 QY 315 KLOLAPNDPPLNTLLKQTLKDTQHYLAHSSDGSCHLLDNKGLHFDIKSTATSYSV 374
 Db 365 KLTLL-----DQOQ--LKOTQGHDTDRWFYSWOYDV 392
 QY 375 LHNSHPGETKGLAQAGTGSVSDGSKGIS-----LGSGTQSHNKTMLSQPGBAH-- 425
 Db 393 TRE-----REQLOQAGS--TVAASGSAKULISTQEDVKLLGANYSADRALSVAARDVHLA 445

QY 426 -----RSLTGIWQH-----PAGARPQGE-----SIRLHDDK 453
 Db 446 GLVEKDKSSRGYQRNHTSSLTGRWNSDESESLKASERSEGETLTKAGRNVSTQCAK 505
 QY 454 IHLHPGLGWQSDAKDTHSOLSRQADGKLYALKDNRTLO---NLSNKSSEKLVDKIS 510
 Db 506 VHAQRD-----LTIDAQNIQVGVQKQYANAKAVRDKTSMGGIGGDNKNS-----552
 QY 511 YSVDRQGVAILTDPGRHMSIMPSLDASPEHSISLSPADAHQGLLHGKSELEAQ--568
 Db 553 -----NREISHASELTSGTLRL-----NGQGVTTITGSKARQKG 589
 QY 569 -SVAISHRLVWADSEKLSAALPKQODGNELKWKAMPQHALDEHGFHDHQSFGFHD 627
 Db 590 GEVTAHGGGLRIIDA---LSTTVDK-----IDARTGTAFNITSSSHK-628
 QY 628 HGQNALVKNFRQOAHACPLNDHQFHPGNLT-----DALVIDNOLGLHHTNPEPHEIL 682
 Db 629 -----ADNSYQSSTASELKS-----TNLTILVSHKQADVIGSOVA-----663
 QY 683 DMGHGLSIALQEGKHLHYFDQITKGWITGAESDCKQLKGLDGAAYLLKDGVEKVRKLNQST 742
 Db 664 -----SCGE-----LSVESKT 674
 QY 743 SSIKHGTENVFSLPVHRNKPPEPGDALQGLNKKDQAQAMAVTGVNKLALTEKGD-----796
 Db 675 GNI-----NVKA-----AERQONIDEOKTALTNGYAK-----EAGDKQYRAG 712
 QY 797 --IRSFQIKPGTQOLERPACTLSREGISGELKDIHVDHKQNLIALTHEGEVFGHPREAWQ 854
 Db 713 LRIEHTRDSEKTRTENSASSLS--GGSVLK-----AEKDVTFSGSKLVAD 757
 QY 855 NGAESSWHKIALPQSSSKLSKSDMSHEKPIATFEDGSQHLKAGGWHAYAAPERGFLA 914
 Db 758 KQDASVGNKVSLAADKTAS-----NTEQTKIGGFY-----792
 QY 915 VGTSGSTVFNRLMQGVKGKVIPOSGLTVKLSAQGTGWTGAERKVSFKSERIRA--Y 971
 Db 793 --TGG-----IDKLGSGVE-----AGYENKTKQAQSSKALTSQSDV 826
 QY 972 AFNPTMTPTPIKNAATATQHQWQREGKPLIYEMQALIKQLDAHNVRHNAPQDLOSK 1031
 Db 827 KGNLTINARDKL-----TOGQAO-----HVGQAY--QENAAQVVDHLAAADTASTT 870
 QY 1032 LETLDLGEH--GAELNDMKRPRDLEOSATRSVTVLQHQGVKLSNGEINSEFPSPGKA 1090
 Db 871 TTKTDVGNIGANV--DYSATVRPERAVGKAAL--DATGVINDIGIGA---PNVG--921
 QY 1091 LVQSFNVRNSQDLSK--SLQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQKEIPLGR 1148
 Db 922 ---LDIGAQGSSEKRSSSQAVVSS-----VQAGSIDINAKGEV---958
 QY 1149 QRPNDKTALTKSLILDTVTIGELHELADKALVSDHKPDADQIKQLRQFPDILREKRY 1208
 Db 959 -RDQGTQYQASK-----GAVNLTAQ-----SHRSEAAANRQDEQSRDT-----R 996
 QY 1209 ESNPVKHYTDMG-----FTHNKALEANYDAVKAFINAFKKE---HHGVN 1249
 Db 997 GSAGRVYVTTGSLITVDKAGEGQGTQRSNSSASQAVTGSIDAANGINNVVKDAIYQCTA 1056
 QY 1250 LT---TRTVLESQSAEL-----AKKLNTLLSLDSG-----BSMSFSRSYGGV---1291
 Db 1057 LNGRGKTAVNAGGDIRLDAQSDKQSESRSGFNFKASAKAGFTADSKNFGAGFGGTHNG 1116
 QY 1292 ----STVFPVTLKKVPVPIPGAGITLDRAYNISFSRTSGGLNVSG-----1335
 Db 1117 ESSSSTAQVGNISGQGVVELKAGRDLTLQG-----TDVKSQGVDSLSAGNKVALQAESTQ 1172
 QY 1336 --RDGVSGNTWATGHVDVPMYTKTKTSAGNASD-----WLS 1371
 Db 1173 TRKESKLSGNDLQAGSSDSKEKTKGNLSAGAFDIAKNESATERQOATIASDKGVTLS 1232
 QY 1372 AKHKISPDLRI--GAAVSG---TLQ-----GTLQNSLKFKLDELPGFTHGLTHTLTPAE 1423

Db 1233 ANKGGDHALHQAKVSGGALEAKNGGILLSEAKNEQHKN-----1275
 QY 1424 LLQKGIHQMKQSKLTFSVDTSANLD-----LRAGINLEDGSKN-GVTA 1469
 Db 1276 -WSLGIKANAKGGO--TFNKDAGKVDPTNGKDTHTLGLAGLKVGEQODKTHANTGITA 1332
 QY 1470 -RVSAGLSASANLAAGSRERSTTSGQFSTTSASNNRPTFLNGV-----GAGA 1516
 Db 1333 GDTVLSNGKDTRLAGARVDADSVQGVGGDLHV--ESRKDVENGVKVDVDAGLSHSNDPGS 1391
 QY 1517 NLTAALGVAIHSSTHEGK-----PVG-----IPPAFTSTNVSAALADNRTSQSISLELKRAE 1568
 Db 1392 SITSKLSKVGTTPRYAGKVEKLEAGVNVKADATTDKYNVSARRLDPOQDITGAVSFSAE 1451
 QY 1569 -----PVTSNDISEL-----TSTLGHKFKDSATTMKLAALKELDDAKPAQLHILOH 1616
 Db 1452 GKVTLPATPAGERKQPGFLWDRGARTVGGAVKDSIT-----GPAGR-----QGH 1494
 QY 1617 FSA-----KDVVGDERYEAVRNLKKLIVIRQQA-----ADSHSMELGSASHSTTVNN 1662
 Db 1495 LKNVADVNNNAVGEQSAIAGKNGVALQVGQOTLTGTGETRSQQQKVELGGSQVSDVN 1554
 QY 1663 LSRINDGIVELLKHFDALPASPASAKRLGEMWMDPALKDIIKLOLSTPSSASVSMEL 1722
 Db 1555 GQRYQGGGRYD-----AAATVGGLLGG--AAKQSV--AGNVPPFASGHASTQQ 1597
 QY 1723 KDGLREQTEKAILDGK 1738
 Db 1598 AD-----AKAGVFSGK 1608

RESULT 6
 TLN1 MOUSE
 ID TLN1 MOUSE STANDARD; PRT: 2541 AA.
 AC P26039; Q8VEF0;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DB Talin 1.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Fibroblast;
 RX MEDLINE=91015390; PubMed=2120593;
 RA Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O.;
 RT "Sequence and domain structure of talin.";
 RL Nature 347:685-689(1990).
 RN [2]
 RP SEQUENCE OF 1603-2541 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.I., Rubin G.M., Hong L.,
 RA Stapleton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Probably involved in connections of major cytoskeletal
 CC structures to the plasma membrane. High molecular weight
 CC cytoskeletal protein concentrated at regions of cell-substratum
 CC contact and, in lymphocytes, at cell-cell contacts.
 CC -!- SUBUNIT: Binds with high affinity to vinculin and with low
 CC affinity to integrins.
 CC -!- PTM: Phosphorylated.
 CC -!- SIMILARITY: Contains 1 FERM domain.
 CC -!- SIMILARITY: Contains 1 I/LWEQ domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X56123; CAA39588.1; -;
 DR EMBL; BC018557; AAH18557.1; -;
 DR PIR; S11661; S11661.
 DR MGD; MGI:1099832; Tln.
 DR GO; GO:0005925; C:focal adhesion; IDA.
 DR InterPro; IPR000299; Band 4.1.
 DR InterPro; IPR002558; ILWEQ.
 DR Pfam; PF00373; Band 4.1; 1.
 DR Pfam; PF01608; I_LWEQ; 1.
 DR ProDom; PD011820; ILWEQ; 1.
 DR SMART; SM00295; B41; 1.
 DR SMART; SM00307; ILWEQ; 1.
 DR PROSITE; PS00660; FERM_1; 1.
 DR PROSITE; PS00661; FERM_2; 1.
 DR PROSITE; PS00057; FERM_3; 1.
 DR PROSITE; PS00945; I_LWEQ; 1.
 DR Structural protein; Cytoskeleton; Phosphorylation.
 KW DOMAIN 86 403
 FT FERM.
 FT DOMAIN 2340 2533 I/LWEQ.
 FT VARIANT 1105 1105 L -> P.
 FT VARIANT 2180 2180 K -> M.
 SQ SEQUENCE 2541 AA; 269832 MW; 14EF75ABE9FC2CBB CRC64;

Query Match 2.3%; Score 219.5; DB 1; Length 2541;
 Best Local Similarity 19.4%; Pred. No. 0.025;
 Matches 444; Conservative 295; Mismatches 860; Indels 693; Gaps 110;

QY 11 KAAVHTAANPVG---HG-VALQ---QGSSSSSPQN-----AAASLAAGKNNKMPR 56
 DB 428 KSTVLQQVNRVGVKVEHGSVALPAMRSGASGFENFQVGSMPPAQQOITSGQMRGHMPP 487
 QY 57 IHQPSTAADG-----ISAAHQKKFSRLRCLG-----TKKFSRPAQOGPGTTHSK- 103
 DB 488 LISAQALGTGTINSSQVQAQAATLDDFTLPLGQDAASKAWRNKNKMDSEKHEIHSQV 547
 QY 104 -----GATLRDLLARDGDTGHE-----AAAPDAARLTR--SGGVK--RRNMDMA--GRP 148
 DB 548 DAITAGTASVNLTAGDPAETDVTAVGCAVTTISNLTMSRGVKKLALLEDGNGRNP 607
 QY 149 MVKGGSGEDKVPFQQKRLHNLNFGMRQTMLSKVAHPASNAGDRLOH-----SPHPIDP 203
 DB 608 LQAARKGLAGAVSELLRSQAASAPRQNLQAGNVGQA--SGELLQOIGESDTPDHFQD 666
 QY 204 SHHEIKKEPVGTSKATTAHADRVETLAQEDDDSEFQQLHQORLARERENPPPKLGVA 263
 DB 667 VLMQL-ANAVASAAALVLKAK--SVAQRTEDSGLO---TQVIAAATQCALSTSLVACT 720
 QY 264 PTISARFQPKLTA-----VAESVLEGTDTTQSPKLPQSKMLKSGAGVTPLAYT 310
 DB 721 KVVA---PTISSVPCQELVEAGRLNAKAVEGCVSQAATEDGQLLRGVGAAT--AVT 775
 QY 311 LDKGKLQALPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLNLDNRKHLFDIKSTAT 370

Db 776 -----QALNELL-----CHVKAHTAGAGPAGR-----YD-QATDT 804
 QY 371 SYSVLHN-----SHPGEIKGK---LQAAGT---GSYVDGKSGKISIGSGTQSHNKTMLS 419
 Db 805 ILTVTNIPISSMGDAGEMVRQARILQAUTDLVNAIKADAE-----GESDLENSKILTS 858
 QY 420 QP-----GEAHSRLTIGIWOHPAGAAPQGHISIRLHDDKIHLHPGLGVQWQADKTHSOLS 476
 Db 859 AAKILADATAKMV-----EAAKGAA-----AHPD-----SEEQQQLR 891
 QY 477 ROADGKLYALKONRTYQNLSDNKSEKLYDKIKSYVDQRGQVAILTDTPRHMSIMPS 536
 Db 892 EAAEG--LRMATNAAQ-----NAIKKLVQRL--HAAKQAASATOTIAAQAAS--APX 943
 QY 537 LDAPESHLSLSLH--PADAHQGLHCKSELEAQSVAISHGRLVVADE-----GKLFS 589
 Db 944 ASAGPQLLVQSCKVAEIPLLVQVGRSQAPSPSAQLALIAASQSFQGGKMWAA 1003
 QY 590 A---IPKQDGNELKMKAMPQHALDEHFHGHQISGFFHDDHGLNALVKN----- 637
 Db 1004 AKASVPTIQD---QASAM-----QLSQCAKNLGTALAE 1035
 QY 638 --NFRQOHAC--PLGNDHQHPGWNL-----TDALVIDNQLGHHNTPPEHIL---D 683
 Db 1036 TAAQKAQEAACGPLEMDSALSVVQNLKDLQEIKAARDGKL-----KPLPGETMEKCTQD 1090
 QY 684 MGH-----LGSALQEGKLYHPDQLTGKWTGAESDCKQLKGLDGAAYLLKDG 731
 Db 1091 LGNSTKAVSSAIAKLGEIA--QGNENYAGIARADVAG--LRSIAQAARGVAALTS 1145
 QY 732 EVKRLINOSTSSIKHGTENVFSLPHVRNKPEPGDALGLNKDDKAQAMAVIGNKYL-- 789
 Db 1146 AVQAVLDTASDVLDRKASSLIEBAKASGHPGPEQORLAQVAKAVTQA---LNRVC 1202
 QY 790 -----ALTEKGD---IRSFQIKPTQQLERPAQTLRSRGISG---ELKDIHV 830
 Db 1203 LPQQRVDNALRAVGDASKRLSLDLPLPPTGTGTQE--AQSRLENAAGLNAQATLVQASR 1261
 QY 831 DHKQNL-YALTHEGEVHFQPREAWQNGAESSSWHKALPQSE-----SKLSLDSHSH 883
 Db 1262 GTPQDLARASGRFGQDFSTFLEA---GVEMAG---QAFQEDRAQVSVNLKISMS 1314
 QY 884 -----KPIATFDGSOHLKAGGWHAAYAAPERGLAVGTSGSQTV----- 923
 Db 1315 LLLAAKALST--DPASPENLKS---QLAAAA--AVTDSINLIITMCTQAPQKQEC 1365
 QY 924 -----FNLMOGVKGVIPGSGLT-VKLSAQTG-----GMT 953
 Db 1366 ALFQLETVRELLNPVQPIINDMSYFGCLDSVMENSKVLGEAMTGISQNAKGNL 1425
 QY 954 GABGRKVSFKFSERIRAYAFNPTMSTPRPIKNAAYATQHWQREGKLPLEYMOGALIK 1013
 Db 1426 IATASALCGFTEAAQAAYLVGVSDP-----NSQACQGLVPTQARA--- 1470
 QY 1014 LDANVVRHNAQPDLOSQKLETLDLGHHG---ABELNDMKRFRDELBQSATRSVTVLG 1070
 Db 1471 -----NQAIQMACQS-----LGEPPGCTQAQVLS-----AATIVAKHT 1502
 QY 1071 GVILKSNGEINS--EFKPSFGKALVQSF-VNVRSGQDLSKSLQ-----QAVHAT 1115
 Db 1503 SALCNSCRLASANTAPTAKRFQVQSAKEVANSTANLVTKIKALDGDGFTTEENRA 1562
 QY 1116 PPSAES--KLQSLMLGH-----FVSAGVDMHQKEIPLGR--- 1148
 Db 1563 APLEAVDNLASFASPESSVPAQISPEGRAAMEPIVISAKTMLESAGSLIQTARALAV 1622
 QY 1149 -QRDPNDKALT--KSLIILDTV----- 1168
 Db 1623 NRPDPWRWSVLGHSRTVSDSIKLITSMRDKAPQGLECETATAALNSCLRDJDQASLAA 1682
 QY 1169 -----TIGELHELADKAKLVSDHKPDADQ---IKOLRQOQFD- 1201

Db 1683 VSQOLAPREGISQBALHTQMTLVAQVISHLIE--PLASAAAEAEASQLGHKVSQMAQYFEP 1740
Qy 1202 -TLREKREYESPVKHYYTDMGFTH-----NKALEANYDAVKAFINAFKKEHGHVNLNLTRT 1254
Db 1741 LTLAAVGNASKTSLHPQOMALLDQTKTLAESALQLLTYTAKAGNPKQAAH-----TQE 1794
Qy 1255 VLES--QCSAEELAKKLNLTLSLDSGESMSFSRSGYGGVSTVFVPTLSKKVPVPIQAG 1312
Db 1795 ALEBAVQMMTEAIVEDLTTTL-----NEAASAAGVVGGMVDGI-TQAINQLDGGPMGDPEG 1848
Qy 1313 ITLDRYNLSFSRSTSGGLNVSF-----GRDGGVSGNIMVATGHDVMPYTGKKT 1362
Db 1849 SFVD--YQTTMVRVAKALAVTQEMVTKNSPPELGPLANQLTSDYGRLASQAKPAAVA 1906
Qy 1363 AGNASDMLSAKHKISPDILRIGAAVSGTQGTQLQNSLKFELTEDELPGFIHGLTHGTLTPA 1422
Db 1907 AENEIEGAHVKHRVQ-ELGHCCSALVTKAGALQCSPSDVYTKEL-----TECA 1954
Qy 1423 ELLOKQIEH---QMKQSKLTFSVDTSAN-----LDLRAGI-----NLNEDGSKP--- 1464
Db 1955 RRVSEKVSHVLAALQAGNRGTQACITTAASAVSGIADLDTTIMEFATAGTLNREGAETPAD 2014
Qy 1465 --NGV--TARVSA-GLSASANLAGSRSRSTSGQ-----FGSTTSASNNR 1505
Db 2015 HREGILTKAKVLVEDTKVLQNAAGSQEKLQAQAQSSVATITRLADVVKLGAALGADDP 2074
Qy 1506 PTFINGVGAGANLTAALGVASHSTH--EGKPVGIFPAFTSTNVSAALALDNRTS----- 1557
Db 2075 ETQVVLINAVKDVAKLGLLISATKAAAGK-VGDDPAVMQKLNKSAKVMVNTVTSLLTKVK 2133
Qy 1558 -----QSTISLEL-----KRAEPTVTSNDISBLTSTLGHFKDSSATTK 1593
Db 2134 AVEDEATKGTALREATTETHEIRQELAVFGSPPEPAKTSPTDFIRMTKGI-----TMATAK 2188
Qy 1594 MLA-----ALXELDKAPAEOLHILQOHFSKADVVGDYERAEVNRMLKLIV----- 1638
Db 2189 AVAAGNSCROEDVITATNLSRRAADMLRACKAEAFHPEVADPVRLRALHYGRECANGYL 2248
Qy 1639 -----IRQAAADSHSMELGSASH-----STTYNNLS 1664
Db 2249 ELLDHVLLTLQKPNPDLKQLTGHSKRAGSVTELIQAAEAMKGTWEDPDPTVIAENE 2308
Qy 1665 RINNDGIVELHGHFDALPASSAKRLGEMMNDPALKDIIKQLQSTPFSFSSVSMELKD 1724
Db 2309 LLGAAAAIEAAAKLEQLKPRAPKEADESLNFEQILEAAKSI-----AAATSALVKA 2362
Qy 1725 GLRROTEKAILDCKVGREVGVLFDNRNLRVKSVSQSYSKS-----EGFN 1772
Db 2363 ASAAQRE-LVAQGVKGAIPANAL--DDGQWSQGLISAARMVAAATNNLCEAANAQVQHA 2419
Qy 1773 TPALLLGTNSAAMSMBERNIGTINFKYQDQDQNTPRFTLEG-GIAQAN-----PQVASA 1825
Db 2420 SQEKLISAKQVAASTAQLLVACKVKADQDSEAMKRLQAAGNAVKRASDNLVKAQAQKAA 2479
Qy 1826 LTDLKKEGLEMK 1837
Db 2480 FEDQENETVVVK 2491

RESULT 7
MLL2 HUMAN
ID MLL2 HUMAN
AC O14686; O14687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein)
DE MLL2 OR ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax.";
RL Oncogene 15:549-560(1997).
RN [2]
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -!- FUNCTION: May be involved in transcriptional regulation.
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBP5, alpha- and beta-tubulins, the trithorax group proteins
CC MLL2 and MLL3, and ASH2/ASCL2.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O14686-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
CC Name=3;
CC IsoId=O14686-3; Sequence=VSP_008560;
CC -!- TISSUE SPECIFICITY: Expressed in most adult tissues, including a
CC variety of hematopoietic cells, with the exception of the liver.
CC -!- MISCELLANEOUS: This gene mapped to a chromosomal region involved
CC in duplications and translocations associated with cancer.
CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 1 SET domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF010403; AAC51734.1; -.
DR EMBL; AF010404; AAC51735.1; -.
DR PIR; T03454; T03454.
DR PIR; T03455; T03455.
DR Genew; HGNC:7133; MLL2.
DR MIM; 602113; -.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007048; P:oncogenesis; TAS.
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR InterPro; IPR003889; FYrich_C.
DR InterPro; IPR003888; FYrich_N.
DR InterPro; IPR000910; HMG_12_box.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR006118; Recombinase.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 5.
DR Pfam; PF00856; SET; 1.


```
QY 785 VNKYLAITEKDIRSFOIKPGTQQLERPAQTLRSREGISGELKDIHVDHKQNLIALTHEGE 844
D 3639 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
QY 845 VFHQPREAWQAGSSSHWIKALQPSBKSLKSLDMSHEHKPIATP-----QQQ 3689
D 3690 QFOQQQQQQQQGMLLNQSLTLLSPQQQQQQQQVAGLPGFMAKPLQHFSSPCALGPTLLTGK 3749
QY 890 -----EDGSHOLKAGWHAYAPERGPPLAVGTSGSTVFNRIMQGVKGV 935
D 3750 BQNTVDPAVSSEATEGPTSHQ-----GGPLAIGTTPS-----MATEPGEV 3790
QY 936 IFG-SG-----LTVKLSAQTGMTGAEGRKVS 963
D 3791 KPSSLGDSQLLVQPOPOPQSSLOLQPPRLPGQQQQQVSLHTAGG--GSHQGLSGS 3848
QY 964 FSEIRAYAFNPTWSTPRPI--KNAAYATQGMQREGKLPKYEMOGALIKOLDANVRH 1021
D 3849 SSE-----ASSVPHLLAQPSVSLGDQPGSMTQNLGPPQPM-----LERPMQNN 3892
QY 1022 NAPONDLQSKLETLDLGHGAELNDMKRFRDELQSATRSVTVLGQHGVLKNGEINS 1081
D 3893 TGPQPKPGLQSGGLPGVIMETVGLRAQL-----QGVLAKNPQLR- 3937
QY 1082 EFKPSPGKALVQSFNVNRSQDLSKLOQAVHATPPSAE-----SKLQSMGLCHFVSAGVD 1136
D 3938 HLSPOQQQQQ--QALLMQRLQ-----SOAVRQTPPYQEPGTQTSPLQGLLGCQPOLGGF 3991
QY 1137 MSHQKGEI-----PLGRQDPNDKATLTKSRILITVTIGELHELADKAKLVSDHKP 1188
D 3992 PGPQTGPIQLGAGPRPGPRLPAPCALSTGVP-----LGPVHTPPSPPOEKKRP 4045
QY 1189 DADQIKQLRQDFDILREKRYESNPVKHYTDMGFTHINKALEANYDAVKAFINAFKKEHGV 1248
D 4046 --SQLPSPSSQLPT-----EAQ-----LPTPTGPT 4068
QY 1249 NLTRTIVLE-----SQGSABELAKLKNLTLLSLDSGESMSFSRSYG-----GGVSTVFPV 1297
D 4069 PKPGQPTLEPPGPRVSPAAQLADTL-----FSKGLGPDWPPDNLAETQKP 4114
QY 1298 TSKKVPVPIPGAGITLDRAYNLFSR-----TSGLNVNFGRDG-----GVSGNI 1344
D 4115 EQSSLVPGHLDVQNGVQVPEASQLSIKQEPREEPCALGAQSVKREANGEPICAPGTSNHL 4174
QY 1345 MVA-----TGHVMPYMTGKT-----SAGNASDWLSAKHKISPDLRIGAAVSGTLQGTQN 1396
D 4175 LLAGPRSEAGHLLOKLLRAKNVOLSTQGSSEGUAEINGHIDSKL--AGLSQKLQGTSPSN 4233
QY 1397 SLKPKLTDELPGFIHGLTHGTLP--AELLQKGIEHQMKQSKLTFVSDTSANLIDLRAGI 1455
D 4234 K-----ED-----AAARKPLTPKPKVQKASDLVSSRKLR----- 4265
QY 1456 NINEDGSKPNQVTA-----VSAGLSASANIAGS--RERSTTSGQFGST 1498
D 4266 --KEDGVRASEALLKQLKQELSLPLTEPAITANPSLFPAPFGSGCPVNGSQLRGAFGS- 4322
QY 1499 TSASNRTFTLNGVAGANLTAALGVAHSSTHEGKPGVIFAFVSTNVNSAALDNRTSQ 1558
D 4323 -----GALPTGPDYVSQLLTKNLSNPPPTPSSLP-----PTP 4355
QY 1559 SISLEKRAEPVTSNDISELSTLKGPKDSATTK--MLAALKELDKAPQOLHIL--QQ 1615
D 4356 PPSVQOKWVNGVTPE-----ELGEHPKDAASARDSEARLDTSEVKSILDLAALPTPP 4409
QY 1616 HFSKDVVDERYEAVRNKKLVTRQQAADSHSMELGSASHSTTVNNLSRINNDGIVELL 1675
D 4410 HNQTEDV-----RMESDESDSP----- 4427
QY 1676 HKHFDAAIPASSAKEL-----GEMMNNDPALKDILIKQLQSTPSSASVSNEL 1722
D 4428 ----DSIVPASPSPESILGEEAPRFPPLGSGRWEQEDRALSPVIPLI-----PRDSIVFPVFPDT 4480
```

```
QY 1723 KDGLREQTEKAILDKGVGREEVGL--FQDRNLRKVSYSVSQSVSKSGFNTPALLLGTS 1781
D 4481 KP-----YCALGLEVPGLPVTTWKKGSEVSVMLTVSAAADKNLNGVMVA 4529
QY 1782 NSAAVSMERNI-----GTINFKYQDQNTPRFTLEGGTAAQNPQVASALTDLKK 1831
D 4530 ELLSMKIPNSYEVLPFPPSPARGGTEPKKGEAEG-----PG-----KE 4567
QY 1832 EGLEMS 1838
D 4568 KGLEKS 4574

RESULT 8
SPOF SCHPO STANDARD; PRT; 1957 AA.
AC Q10411; QOUSE9;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
CN SP015 OR SPACIF3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_taxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20107136; PubMed=10639340;
RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Spol5p is localized
RT to the spindle pole body and essential for its modification.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Squares S., Stevens K.,
RA Skelton J., Simmonds M., Squares R., Squares S., Sharp S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymouprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesti D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu B., Dreano S., Gloux S., Lelaue V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
[3]
RN SEQUENCE OF 705-871 FROM N.A.
RP STRAIN=968 h90;
RC MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
```

CC -!- FUNCTION: Has a role in the initiation of spore membrane
CC formation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Spindle pole body.
CC -!- SIMILARITY: Belongs to the MPC70 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z70690; CAA94624.1; --
CC DR EMBL; AB027811; BAA87115.1; --
CC DR PIR; T38077; T38077.
CC DR GeneDB_Spombe; SPAC1F3.06c; --
CC KW Sporulation; Coiled coil.
CC
CC FT DOMAIN 199 785 COILED COIL (POTENTIAL).
CC FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
CC FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
CC FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
Query Match 2.3%; Score 217.5; DB 1; Length 1957;
Best Local Similarity 18.7%; Pred. No. 0.021;
Matches 386; Conservative 298; Mismatches 731; Indels 645; Gaps 93;
QY 33 SSSPQAAASLAAGKRNKMPRIHQPSST---AADGISAAHQOKKSFSLRGCLGTYKFS 89
DB 18 SASLSVSSAASPFIDSLTPRENIASRTAGQAAEDGDTSSQHEDSSEELK----- 68
QY 90 RSPAQOPGPTTHSKGATRLDARDGCTQHEAAAPDAARLT-RSGGVKRRNMDMAGRP 148
DB 69 ---RQVRGMRHSDLSIDAKLSSEGST---ASSALFLTPRS-----PSNASWL 112
QY 149 MYVGGSGEDKV---PTQKRHQIINFGQRMOTMLSKMAHPASAGADRLQHSPPHIPGS 204
DB 113 LVRGGLDSDPILDINSVTQSNLLNELKQVR---SKLAALHENGILSLQ-----LSSS 163
QY 205 HHEIKKEPVGTSKANTAHADRVIEAQEDDDSDFOQ---LHQORLARERENPPQPKLG 260
DB 164 NKDKK-----NTSSVT-----LTSEEDVSVFQKLTWNESNFSASQSEAYDSLKQL 211
QY 261 VATPISARFQPKLTAAVESVLEGTDTTQSPKQSMKKGSGAGVTPLAVTLDKGKQLAP 320
DB 212 TVTEKDKKEDYKEDYSSIKASLAERQASNKSLRGQERLEKLLVSNK----- 264
QY 321 DNPPALNTLLKQTLG-----KTOHYLAHHAASDGSQHILLDNKGLHFDIKSTATSYS- 373
DB 265 ---TVST-LRQTENSLRAECKTLQEKLEKCAINEEDSKL-----LEELKHVANYSD 312
QY 374 -VLHNSHPGKIGKLAQAGTGSVDGSKISLGSQTQSHNKTMLSQPGEAHRSLLTGI 432
DB 313 AIHV-----KDKL-----IEDLSTRIFEFDNLKSERDT-LSIKNEKLEKL--- 352
QY 433 WQHPAGAAPQGES-IRLHDDKTHILHPELGVQWQADKTHSQLSROADKLYAL-KDNR 490
DB 353 -RTIGSLKDSRTNSQLEEMVEL-----KESNRTHSQLT-DAESKLSFEGENK 402
QY 491 TLQNLSDN-----KSSKELVDKTKSVSDQGVQVAILTDTTPGRHKMSIMPSLDASPESHI 545
DB 403 SLKSGIDEYQNNLSSKDKMKVQVSSQLEEARSLAHAT---GK----- 442
QY 546 SLSLHFADAHQGLLHGKSELEAQSVAISHGLRVVADSEGLFSAAIKPGQDGNELKMKKA- 604
DB 443 -----LAEINSEKDFQNKTKDPEKIEQDLRACINSSS-----NELKEKSA 483
QY 605 -----MPQHALDEHFHQHDIQISGFFHDDHQLNALVKN--NFRQOACPLGNDHQHPG 656
DB 484 LIDKQOELANLREQIKEQKVS---ESTQSSLSQSLQDILNEKKHIE----- 528

QY 657 WNLTDALVIDNQI-----GLHHTNPPEPHEILDMGHLS-----LALQBGKLUHYFDQLTKGW 707
DB 529 -----VYESQNLKELGELQT-----EISNSELSSQLSTLAAEKEAAVATNELSESK 576
QY 708 TGAESDCKQLKXGLDGAAYLLKDG-----VKRLNINQS-----TSSIKHGT 749
DB 577 NSIQTICNAPQOEKLAWSVQMKENQNFSSLDTSFKKLNESHQELNHNHQTITKQLKDT 636
QY 750 ENVFSLPHVR-NKPEPCDALQGLNKDDKAQAMAVIGVKNKYALTEKGDIRSFO--IKPGT 806
DB 637 SKLQQLQLERANFEQKESTLSDENNDLRTKLLKLEESNKS-ILKKQEDVDSLEKNIQTLK 695
QY 807 QQLERPAQTL-----SREGISGELKDIH----- 829
DB 696 EDLRKSEEARLRFSLKAKNLREVIDNLKKGKHETLEAQRNDLHSLSDAKNTNAILSELT 755
QY 830 ---VDHKQ---NLYALTHEGEVPHQPREAWQNGAES--WHKL-----ALPQ 869
DB 756 KSSSEDVKRLTANVETITQDSKAMQSFSTLVNYSQISLNLYHELRDHHVMQSQNTLLE 815
QY 870 SESKLSLDMSHHEHPITAFEDGSQHLKAGGWHAYAAPE-----GKGVK-IPSGLTVKL 945
DB 816 SESKLT-DCENLTQQNMTLIDNVQKLM-----HKHVQESKVSSELKEVNGKLSLDLKNL 869
QY 910 RGPLAVGTSGSQTVFNRIMQ-----GVKGV-----IPSGLTVKL 945
DB 870 RSSLNVAISDNDQILTQLAELSKNYDSLEQESQALNSGLKSLAEKQKLLHTEBELHRL 929
QY 946 SAQTGQMTGAE-----GRKVSXKFSERIRAYAFNPTMTSPRIKNAAYATQHCWQREG 999
DB 930 DKUTGKLTKEESKSDLGKLTARQEE----- 956
QY 1000 LKPLYEMOGALIKQLDAHNRHNAPOFDLQSKL-ETIDLGEHGAELNDMKRFRDELEQS 1058
DB 957 -----ISNLKEENMSQCAITSVKSLDETL---SKSKLEADIEHLKNKVS 1002
QY 1059 ATRSVTVLGOHQGV---LKSNGE-----INSEKPSPKCA 1090
DB 1003 EVERNALLASNERLMDLKNNGENIASLQTEIEKRAENDDLQSLKSVSEYE---NL 1058
QY 1091 LVQSENVNRSQDLSLQQAQVHATPPSAESKLSQMLGHFVSAGV---DMSHOKGEPLG 1147
DB 1059 LLTSSQTNKSLDKTQVLY-----TEKNVQKLLDEKQDNVELELTQVSKL--- 1107
QY 1148 RQRDPNDKALTAKSRILDTVTIGELHELADKAKLYSDHKPDADQIKQLRQOFTLREKR 1207
DB 1108 -----GEENAQIKDELLALRKKSKQHDLC--ANFVDDLKESDLEALQLTNEKNELIVSL 1160
QY 1208 YESNPVKHYTDMGFTHNKALEANYDAKAFINAFK-----EHHGVNLTFTVLESQSAE 1263
DB 1161 EQSN-----SNNEALVEERSDLANRLSDMKKSLSDSDNIVSVIRSDLVVRNDELD 1210
QY 1264 LAKKLNKTLIS-----LDS--GESMSFSRSYGGGVSTVFVPTLSK-KVPVPI 1308
DB 1211 TLKKDKDSLSTQVSEVCQRDDLLDSLKGCEESFNKY---AVSLRELCTKSELDVVPSEI 1267
QY 1309 -----PGAGITLDRAYNLSFRTSGGLN-VSFG- -DGGVSGNIMVATGHVMPYM 1356
DB 1268 LDDNFVFNAGNFSLSRLTVLSLENVLDAFNQVNFKKWELD---NRLTTTDAEFTKV 1322
QY 1357 TGKKTSGAGNASDWLSAKHKISPLRIGAAVSGTIGTLQNLKSL-FKLTDELPGFIHGLT 1415
DB 1323 ADLEKLQHEHDDWLIQR-----GDLEKALDKSEKFLRKEAEMTENTHSLE 1368
QY 1416 HG---TLTPAELLQKGIHQMKQSKLTFSDVTSANLRLRAGLNLEDGSKPNGVTVARVS 1472
DB 1369 EGKEETKKEITAEISSRLLEDNQLATNKLKNQD-----HLNQETRLKEDVLK-----EKES 1418
QY 1473 AGLSASANLAAGSRERSTTSGQFSTTSANNRPTFLNGVG--AGANLTAALGAHVASTHE 1531
DB 1419 LIISLEESL---SNQRKES---SLLDKAKNELEHMLDSTSRKNSSLMEXIESINSSLDD 1471
QY 1532 GKPVGIFPAPTSTNVSAAL-----ALDNRTSQSISLELKRAEPTVTSNDISLTSLTGKHK 1587

RL Nature 409:529-533 (2001).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RMD 050952;
 RC MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -!- SIMILARITY: Contains 16 Big-1 domains.
 CC -!- SIMILARITY: Belongs to the intimin/invasin family.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 1315.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB005423; AAG57041.1; -;
 DR EMBL; AP002559; BAB36198.1; ALT_FRAME.
 DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
 DR InterPro; IPR003344; Big_1;
 DR InterPro; IPR003535; Intimin.
 DR InterPro; IPR008964; Invasin_intimin.
 DR Pfam; PF02369; Big_1; 16.
 DR PRINTS; PR01369; INTIMIN.
 DR SMART; SM00634; BID_1; 16.
 DR SMART; SM00889; PKD; 8.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834 BIG-1 1.
 FT DOMAIN 840 929 BIG-1 2.
 FT DOMAIN 931 1033 BIG-1 3.
 FT DOMAIN 1042 1132 BIG-1 4.
 FT DOMAIN 1134 1236 BIG-1 5.
 FT DOMAIN 1245 1335 BIG-1 6.
 FT DOMAIN 1337 1439 BIG-1 7.
 FT DOMAIN 1448 1539 BIG-1 8.
 FT DOMAIN 1548 1652 BIG-1 9.
 FT DOMAIN 1653 1750 BIG-1 10.
 FT DOMAIN 1751 1855 BIG-1 11.
 FT DOMAIN 1856 1957 BIG-1 12.
 FT DOMAIN 1963 2056 BIG-1 13.
 FT DOMAIN 2065 2156 BIG-1 14.
 FT DOMAIN 2157 2252 BIG-1 15.
 FT DOMAIN 2254 2355 BIG-1 16.
 SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;
 Query Match 2.2%; Score 205; DB 1; Length 2660;
 Best Local Similarity 19.2%; Pred. No. 0.13;
 Matches 379; Conservative 246; Mismatches 775; Indels 578; Gaps 92;
 QY 170 NFGQMRQTML-----SKMAHPASANAGDRLOHSPHPIPGSHHEIKKEEPVGTGSK 218
 DB 514 NFNREQSMVVVQAPTLSQKDSVSLSSQTLSD--SHSTATLTFIAHDAAGNEVIGL-V 570
 QY 219 ATTAHADRVEIAQED----DSEFQQLHQQLARERENPPQ-----PPKLGVATP 264
 DB 571 LSTRHEGVQDITLSDWKDNGDGTQTLTTGAMSGTTLTLPQNGVDAAKAPAVNNIISV 630
 QY 265 ISARFPQKLTAVASVLEGT-----DTQSPKLPQSMKSGGAGVTPPLATLDKKG 315
 DB 631 SSSRTHSSIKDKRYLSGNPIEVTVELRDENDKPVKEQKQLAT-----AVSIDNVK 683
 QY 316 LQLAPDNPALNTLLKQTL-OKOTQHYLAHASSDGGQHLLLDN-----KGLHFDIKS 367

DB 684 -----PGVTTDWKETADGVYKATYATYTKGSLTAKLLMQNWNEDLHTAGFIIDANP 735
 QY 368 TATSYSVLHNSHPGEBIKKLAQAAGTGSVSV--DGKSGKIS-----LGSGTQSHN--- 414
 DB 736 QSAKIATLSASNGVLANENA-ANTSVNVADDEGSPINDHTVTFVAVLSSGATSFNQNT 794
 QY 415 -KTMLSQFGEAHRSLLTIGWHPAGAARPOGESIRLHDDKIHLHPGLGVWQSA----- 467
 DB 795 AKTDVN-----GLATFDLKSSKQEDNTEVT-LENGVQKTLIVSVFG 835
 QY 468 -----DKDTHSQLSQAQGLYALKNDRTLQNLNKNSEKSLVD 506
 DB 836 DSSTAQVDLQSKNEVVADGNSATMTATVRDAKGNL--LNDVKVTFNVNAAA----- 887
 QY 507 KIKSYVDQGOVAILITDTPGRH-KMSIMPSLDASPESHISLSLHFADAHQGLLHGKSEL 565
 DB 888 KLSQTEVNSHDGIATATLSLKNGDYVTASVSSGSOAN-----QQVIFIGDOST 937
 QY 566 EAQSVAIASHGRVLVADSEGKLFSAAIKQGDGNEELKMKAMPQHALDEHFGHDHQTSGFFH 625
 DB 938 AALTLSPVSGDITVTNT-APLHMTATLQDKNGNPLKQKEIT-----FSVPNDVASRFS 989
 QY 626 -----DDHGQLNALVKNNFRQOH--ACPLGNHQHFHPCWMLTD-----ALVIDNOLG 670
 DB 990 ISNSGKGMTDSNGTAIASLTGTLAGTHMITARLANS-----NVSDTPQMTFVADKDR 1042
 QY 671 LHHTNPEPHEITLDMHGLSLAQGLKHLHYFDOLTKWGTG----- 709
 DB 1043 VVVLQTSKAEIING-VDEYTLTATVKDPDNNVKNLSVVPRTSPATQLSLNARNTNEN 1101
 QY 710 --AESDCKQLKGLDGAAYLLKDG--EYKRLININQSTSS-----IKHGTENVF 753
 DB 1102 GIAEVTLKGTVLGVHTAEAILLNGNRDTKI VNIAPDASNAQVTLNIPAAQVVNNNSDSVQ 1161
 QY 754 SLPHVR--NKPEPGDALQGLNKDDKA-----QAMAVIGVKNYKALTEKGDIRSFOIK 803
 DB 1162 LTATVKDPSNHPVAGITVNFPTMPQDVAANFTLENNGIAITQANGEAHVTLKGG--K 1215
 QY 804 PGTQOLERPACTLSREGISGELKDIHVDHKQNLAL--THEGEVHFHQPREAWQNGAESSS 861
 DB 1216 AGTHTV---TATLGNNNASDAQVTFVADKDSAVVVLQTSKAEII-----GNGVDETT 1265
 QY 862 WHKLALPQSESKLKLSDMSHEHKPIATFEDGSQHLKAGG-----WHAVAAP 908
 DB 1266 LTATVKDPSNHPVAGITVNFPTMPQDVAANFTLENNGIAITQANGEAHVTLKGGAGTHVTATL 1325
 QY 909 ERG-----PLAVGTSGSQTVFNRLMQGVKGVIPGSLGLTVKLSAQGTGGMGTGAEGRKVS 961
 DB 1326 PNGNNDTKI VNIAPDASNAQVTLNIPAQ---QVVTNNSDSVQLTA-----TVK 1370
 QY 962 SKFSERTRAVAFNPTMTSTPRPI-----KNAAVATQHGMOGREGKLPXYEMOGALIKQL 1014
 DB 1371 DPSNHPVAGITVNFPTMPQDVAANFTLENNGIAITQANGEAHVTLKGGAGTHVTATL 1428
 QY 1015 DAHNVRHNAFPQDLQSKLETIDLGEHGAELNMDKFRDELE---QSATRSVTVLGQHQ 1070
 DB 1429 SNNNTSDSQPVTFFVADKTSALV-----LQISKNEITGNGVDSATLITATVKDQF- 1477
 QY 1071 GVLSKNGEINSEFKPSGPKALVQSFNVNRSGDLSKSLQQA VHAHATPPSABSKLQSMGLGHF 1130
 DB 1478 -----DNEVNN-----LPVTFSTASSGLTLT-----PESNTNESGIAQA 1512
 QY 1131 VSAGVDMSHQGEIPLGRQRPDNDKATLTKSRLITDVTIGELHELADKAKL-VSDHKPD 1189
 DB 1513 TLGAVAFGEQTVTASLANNGASDNKT-----VHFIGDTAAAKIETLTPV 1556
 QY 1190 ADQI---KQLRQOPDTLLREKRYESN--PVKHYTDMGFTHNKALEANYDAVKAFINAPKKE 1244
 DB 1557 PDSIIAGTPQNSGSSVITATVVDNNGFPVKGVTVNFTSNAATAEMTNGGCAVTN---E 1611
 QY 1245 HHGVNLT---TRTVLES-----OGSABELAKKLK-----NTLL-----SL 1275
 DB 1612 QGKATVYTTNTRSSIESGARPDTVASLENGSSLTSTSINVNADASTAHLTLLQALPDTV 1671

CC IsoId=P55937-2; Sequence=VSP_007730;
 CC Note-No experimental confirmation available;
 CC TISSUE SPECIFICITY: Highly expressed in testis. Transcripts can be
 CC found in spermatids during spermatogenesis. No expression in
 CC Leydig cells, spermatogonia or spermatocytes. Detected at low
 CC levels in all tissues.
 CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.
 CC -!- PTM: Cleaved by caspases in apoptotic cells (By similarity).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D78270; BAA19612.2; -;
 CC EMBL; AB029537; BAA86889.2; -;
 CC EMBL; AB029531; BAA86889.2; JOINED.
 CC EMBL; AB029521; BAA86889.2; JOINED.
 CC EMBL; AB029522; BAA86889.2; JOINED.
 CC EMBL; AB029523; BAA86889.2; JOINED.
 CC EMBL; AB029524; BAA86889.2; JOINED.
 CC EMBL; AB029525; BAA86889.2; JOINED.
 CC EMBL; AB029526; BAA86889.2; JOINED.
 CC EMBL; AB029527; BAA86889.2; JOINED.
 CC EMBL; AB029528; BAA86889.2; JOINED.
 CC EMBL; AB029529; BAA86889.2; JOINED.
 CC EMBL; AB029530; BAA86889.2; JOINED.
 CC EMBL; AB029531; BAA86889.2; JOINED.
 CC EMBL; AB029532; BAA86889.2; JOINED.
 CC EMBL; AB029533; BAA86889.2; JOINED.
 CC EMBL; AB029534; BAA86889.2; JOINED.
 CC EMBL; AB029535; BAA86889.2; JOINED.
 CC EMBL; AB029536; BAA86889.2; JOINED.
 CC EMBL; AB029537; BAA86890.2; -;
 CC EMBL; AB029521; BAA86890.2; JOINED.
 CC EMBL; AB029522; BAA86890.2; JOINED.
 CC EMBL; AB029523; BAA86890.2; JOINED.
 CC EMBL; AB029524; BAA86890.2; JOINED.
 CC EMBL; AB029525; BAA86890.2; JOINED.
 CC EMBL; AB029526; BAA86890.2; JOINED.
 CC EMBL; AB029527; BAA86890.2; JOINED.
 CC EMBL; AB029528; BAA86890.2; JOINED.
 CC EMBL; AB029529; BAA86890.2; JOINED.
 CC EMBL; AB029530; BAA86890.2; JOINED.
 CC EMBL; AB029531; BAA86890.2; JOINED.
 CC EMBL; AB029532; BAA86890.2; JOINED.
 CC EMBL; AB029533; BAA86890.2; JOINED.
 CC EMBL; AB029534; BAA86890.2; JOINED.
 CC EMBL; AB029535; BAA86890.2; JOINED.
 CC EMBL; AB029536; BAA86890.2; JOINED.
 CC EMBL; BC043452; BAA86890.2; -;
 CC EMBL; BC053002; BAA86890.2; -;
 CC EMBL; AK032610; BAC27949.1; -;
 CC EMBL; T42722; T42722;
 CC MGD; MGI:96958; Golga3.
 CC GO; GO:0005793; C:ER-Golgi intermediate compartment; IDA.
 CC GO; GO:0000139; C:Golgi membrane; IDA.
 CC GO; GO:0005515; F:protein binding; IPI.
 CC InterPro; IPR003345; M_repeat.
 CC Pfam; PF023370; M: 5.
 CC Antigen; Spermatogenesis; Developmental protein; Golgi stack;
 CC Coiled coil; Phosphorylation; Alternative splicing.
 CC FT DOMAIN 132 217 GOLGI-TARGETING DOMAIN (BY SIMILARITY).
 CC SER-RICH.
 CC POLY-ALA.
 CC COILED COIL (POTENTIAL).
 CC GLN-RICH.
 CC GLN-RICH.
 CC G -> GVDGFHDNLRNSQGTSAEGSVRKEALQSLRLSLPMQ

FT ETOLC (in isoform 2).
 FT /FTID=VSP_007730.
 FT T -> I (IN REF. 3; AAH53002).
 FT CONFLICT 30 30
 FT CONFLICT 437 437 G -> E (IN REF. 3; AAH53002).
 FT CONFLICT 552 552 M -> T (IN REF. 2).
 FT CONFLICT 803 803 Q -> L (IN REF. 2).
 FT CONFLICT 942 942 T -> S (IN REF. 3; AAH53002).
 FT CONFLICT 975 975 A -> S (IN REF. 3; AAH43452).
 FT CONFLICT 1426 1426 A -> P (IN REF. 4).
 FT CONFLICT 1447 AA; 162820 MW; 88B68F2D142F8A9B CRC64;
 FT SQ SEQUENCE 1447 AA; 162820 MW; 88B68F2D142F8A9B CRC64;
 Query Match 2.2%; Score 204; DB 1; Length 1447;
 Best Local Similarity 18.9%; Pred. No. 0.057;
 Matches 316; Conservative 257; Mismatches 675; Indels 426; Gaps 77;
 QY 112 ARDDG----ETQHEAAAAPDAARLTRSGGVKRNMDMDAGRPVKGSG--EDKVPTQOK 164
 DB 6 AKDGLWESKSSDDVSCPEASLETVGSRLRPDQDQAQDASVENRKGESGSPDRSS 65
 QY 165 RHQNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHPPGSHHEIKEEPPVGSSTKATTAHA 224
 DB 66 QVAICQNGQIPDLQLS--LDPTTSPVG-----PDASTGSTASSLPLEK 106
 QY 225 DRVEIAQEDDDSEFOQLHQORLARENPPOPPLGVATPISARFQPKLTAVAESVLEGT 284
 DB 107 EEQVRLQARKLE-EQLMQYRVKRHERSSQP---ATQMKLFSTLDPELM----- 152
 QY 285 DTTQSPKPKQSMKLGSGAGVTPLAVTLDKGLQLAPDNPALNTL-----LKQT 333
 DB 153 -----LNFENLPRAS-----TVAVTKEYSFLTSVPRGPKVSLGSLLAHSKKKSKSS 201
 QY 334 LKQDTQHYLAHASSDGSQHLLDNKGHLFDIKSTATSYSLVNSHPGIEKGLAQAGTG 393
 DB 202 KIRSLADYRTEDPDSGLGSTADAVGS--SLQSRSTSVVSEVSPSSTDNRVE---- 255
 QY 394 SVSVGKSKGKISLGSTQSHNKTMLS-----QPGEAHRSLTGT-----IWQ 434
 DB 256 SASMTGDSVSEADGNESDSSSHSLSARGACGVLGVMGPGTAY--MVDQETSAEALGQ 313
 QY 435 HP-----AGAAPQGES-----IRLHDDKTHILPELGVWQSDKQTHSOLSRQADG 481
 DB 314 FPSIKQVLOAAAQHQDQOEANGEVRSRDSI-----CSSVSMESSLAEQDE 362
 QY 482 KLYALKDNRTQLNSDN---KSSEKLVDKIKSYSDPQRGQVAILT---DTPGRHKMSIMP 535
 DB 363 LLQILKDKRLEQVEALSLEASQALQEK-----AELQAQALALSTRLOAQVEHSHSQ 417
 QY 536 SLDASPESHISLHFADAHQGLLHGKSELEAO--SVAISHGLRVADSEGLKFLSAAIPK 593
 DB 418 QDLSLSEVDTLKQSCWDLGRAMTDLQSMLEAKNASLASSNNDLQVAEEQYQRLMA---- 473
 QY 594 QGDGNEKMKAMPQHALDEHFGHDHOLISGFPHDGHQNALVKNPNFQQACPLGNHDHF 653
 DB 474 -----KVEDMORNILSK-----DNIV-----HDLRQMTALQSL----- 503
 QY 654 HPGWNLTDAVIDNQLGLHHTNPEPHEILDMDHGLSLALQEGKLYFDQLTKMGTAESD 713
 DB 504 -----QQVQLERTTLTSK--LQASQAEITSLQHAQWYQQQLT----- 539
 QY 714 CKOLKGLDGAAYLLKDGVEKRLNINQSTSS--IKH-GTENVFSLPH-----VRNK 761
 DB 540 -----LAQEARVLQGEAHIQVQMTQAGLLEHLKLENV-SLSHQLTETQHRSIKEK 591
 QY 762 PEPGDALQGLNKDDKACAMAVIGNVKNLYALTEKGDIRSFOIKPG--TQOLERPAAITLSREG 820
 DB 592 ERITAVOLQSTEADMLDQEAAFVOIREAKTWEEDLQRRLEEFEGEREQKQVADAAA--S 649
 QY 821 ISGELDKIHV---DHKQNLVYALTHEG-EVPHQ---PREAWQNGAESSSMHKALPOSESK 873
 DB 650 LEQOLEQVKLTLFQRDQQLAALQOEHLVDVTKLTSTQEAALQAKQSLDLDLHTRYDELQAR 709
 QY 874 LKSLDMSHEHKPTAT-FEDGSGHQLKAGGHWAAPE---RGLPVGTSQSVTFNRLMQ 929
 DB 874 LKSLDMSHEHKPTAT-FEDGSGHQLKAGGHWAAPE---RGLPVGTSQSVTFNRLMQ 929


```
Db 710 LEEQREADSDRAHFLQNEKIVLEVALQSAKDEKELDRGARRLEEDTEET--SGLE 767
Qy 930 GYVKGVIPGSLTVKLSAOTGGMTGAGKVKSSK-SERIRAVAFNPTWSTPRIPKNAAY 988
Db 768 QIRODLAVKSNQVEHJQQTATILR-KOMQKVEQFVQKQWVEAYRRDATSKQDLNELK 826
Qy 989 ATQHGQWG--REGLKPLYEQG-----ALIKO-----LDAHVNRHNAPO 1026
Db 827 ATKRLDSEMKELRQBELIKQGBKKTVEVHESRLQKDMSLVHQMAELEGHLQSVQKRD 886
Qy 1027 DLQSKLETLDLGHGABLND-----MKFRDELEQSATRSVTVLQHQGVKNSGENE 1082
Db 887 EMEIHLQSLKFKQEQMIALTEANETLKQIEELQEQAKKAIT--EQQKQMKLUGSLTSA 944
Qy 1083 FK--PSPGKALVQSFNVNRSQDLSKSLQQAQVHATPPSAESKLSQMLGHFVSAGVD-MSH 1139
Db 945 QEMKTKHAYENAVSI-----LSRLQEL-ASKEATDAELNQLRAQSTGSSDPVLH 997
Qy 1140 QK-GEIPLGRQRPNDKTKLSRLILDTVTIGELHDLAKKLVSQDHRKEDADQIKQLRQ 1198
Db 998 EKTRALEVELQNVGSKILLELQEBVITWTSQELSEKVELELED--ELQESRGFR 1054
Qy 1199 QDTLREKYENPVKHYTDMGFTHNK--ALEANYDAVK-----AFINAPKKEHGV---- 1248
Db 1055 KIKRLEES-----NKLALALEHERGKLTGLQGSNAALRHNLSILETA 1097
Qy 1249 -----NLTRTVLESQGSABLAKKLNTLLSDSGESMSFSRSGGVSTVFP 1297
Db 1098 LAKREADIYQNLQVAVI--QRKEEDROMQLVQALQ-----VS 1136
Qy 1298 TISKVVPVPIPGATITLDRAVNLFSRTSGGLNVSFGDRGGVGNIMVATGHDVMPYMT 1357
Db 1137 LEKEMEV-----NSLKEQMA 1152
Qy 1358 GKTSAGNASDWLSAKHKISPDLRIGAAVSGTILQ-----TLQNSLKFKLTED-----E 1406
Db 1153 ARIEAGH-----NRHFKAATLEL-SEVKELQAKEHLVQTQABVDELOIQDKHSHOE 1206
Qy 1407 LPQFIHGLHTGLTPAELLQKIEHQWQ--GSK-----LTFSDVTSAN--LDLRAGINL 1457
Db 1207 IAQFQTELAER-TQQLLQKLEQMSQPTGSEMEDLKWELQKEREISLQKQDL 1265
Qy 1458 NE-DGSKPQNGTVARVAGSASANLAAGSRERSTTSQF---GSTTSAGNNRPTFLNGVG 1513
Db 1266 TEQQKKELEGTOQTTLTIKSELEMYQEDLS-ETQDKFEMLOAKVSELKNNMKTLLQ--- 1321
Qy 1514 AGANLTAALGVAHSSTHEGKPVCFIPFTSNVSAALAL---DNRTSQSISLELKAEPV 1570
Db 1322 --QNQQLKDLRRGAKKKEPKG-----BSNSSPATPIKIPCPVPASLLELLRPPFA 1374
Qy 1571 TSND-ISELTSTIG--KHFKDSATTKN-----LAALKELDDAKPAEQH 1611
Db 1375 VSKEPLKNLNNCLQQLKQEMDSIQRQWBEHTITVHESLSWAGV-BAAPAEHAH 1427

RESULT 12
APC MOUSE
ID APC MOUSE STANDARD; PRT; 2845 AA.
AC Q61315; Q62044;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein) (mAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
RC STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;
RX MEDLINE=92263101; PubMed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
```

```
RA Luongo C., Gould K.A., Dove W.F.;
RT "Multiple intestinal neoplasia caused by a mutation in the murine
RL homolog of the APC gene.";
RN Science 256:668-670(1992).
RP ERRATUM.
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RL Science 256:1114-1114(1992).
RN [3]
RP SEQUENCE OF 1-45 FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Dicker F., Lambertz S., Reitmaier A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RN region segments.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RP ALTERNATIVE SPLICING.
RX MEDLINE=94061824; PubMed=8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RT "APC gene messenger RNA: novel isoforms that lack exon 7.";
RN Cancer Res. 53:5589-5591(1993).
CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
CC and participates in Wnt signaling. APC activity is correlated with
CC its phosphorylation state [by similarity].
CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
CC axin [by similarity].
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q61315-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q61315-2; Sequence=VSP_004116;
CC Name=3;
CC IsoId=Q61315-3; Sequence=VSP_004117;
CC Name=4;
CC IsoId=Q61315-4; Sequence=VSP_004116, VSP_004117;
CC -!- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
CC lung, brain, stomach, intestine, testis and ovary.
CC -!- PTM: Phosphorylated by GSK3B [by similarity].
CC -!- SIMILARITY: Contains 7 ARM repeats.
CC -----
CC This SWISS-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; M88127; AAB59632.1; -
DB EMBL; U02937; AAA03443.1; -
DB PIR; I49505; I49505.
DB HSP; Q02248; 3BCT.
DB MGD; MGI:88039; Apc.
DB GO; GO:0005737; C:cytoplasm; IDA.
DB GO; GO:0005634; C:nucleus; IDA.
DB GO; GO:0008013; P:beta-catenin binding; IDA.
DB GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
DB GO; GO:0009798; P:axis specification; IMP.
DB GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
DB GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
DB InterPro; IPR008938; ARM.
DB InterPro; IPR000225; Armadillo.
DB Pfam; PF00514; Armadillo_seg. 4.
DB SMART; SM00185; ARM; 5_seg.
DB PROSITE; PS0176; ARM_REPEAT; 1.
KW Wnt signaling pathway; Anti-oncogene; Phosphorylation;
KW Alternative splicing; Repeat; Coiled coil.
FT DOMAIN 1 61 COILED COIL (POTENTIAL).
FT DOMAIN 125 245 COILED COIL (POTENTIAL).
FT DOMAIN 1 728 LEU-RICH.
FT REPEAT 451 493 ARM 1.
```

```
FT REPEAT 503 545 ARM 2.
FT REPEAT 546 589 ARM 3.
FT REPEAT 590 636 ARM 4.
FT REPEAT 637 681 ARM 5.
FT REPEAT 682 723 ARM 6.
FT REPEAT 724 765 ARM 7.
FT DOMAIN 739 2834 SER-RICH.
FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1864 1891 HIGHLY CHARGED.
FT VARSPLIC 243 276 Missing (in isoform 2 and isoform 4).
FT VARSPLIC 310 410 /FtId=VSP_004116.
FT VARIANT 120 120 T -> A (IN STRAIN CAST/EI).
FT VARIANT 493 493 Y -> I (IN STRAIN CAST/EI).
FT VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).
FT VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).
FT VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).
FT VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).
FT VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).
FT VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).
FT VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).
SQ SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;

Query Match 2.2%; Score 204; DB 1; Length 2845;
Best Local Similarity 19.9%; Pred. No. 0.17;
Matches 412; Conservative 272; Mismatches 808; Indels 580; Gaps 97;

QY 4 KSLGTEHKAHVTAHNPVGHVALQGGSSSSPQNAASLAABGKRGKMPRIHQPSTA 63
DB 242 RSSQSRHDAASHEAGRQHEGHAESNTAASSQSPAT-----RYDHETAS 288
QY 64 ADGISAHQKKSFLRGCLGK---KFSRSAPQGPGTTHSKGATRLDLARDGGETQH 120
DB 289 VLSSGTHSAPR--RLTSHLGTKVEMYSLMLG---THDKDMGRTLLAM----- 335
QY 121 EAAAPDARLTSGGVKRRNDDMAGRPM---VKGSGGEDKVPTQKRHLQNNFGMRQT 177
DB 336 -SSQSDSCISMRSQSGC-----LPLLIQLLHGNDKDSVLLGNSRG----- 373
QY 178 MLSKMAHPASAGADRILOHPPHPIGSHHEIKKEPVGSTSKATTAHADRVIEAIEDDSE 237
DB 374 --SKEARARASAAALHNIHSQDDKGRREIR-----VLHLEQIRAYCETCWE 420
QY 238 FOQLHQRLARENPPOPKLGAVTPI SARFQPKLTAVAESVLEGTDTTQSPKQSM 297
DB 421 WQEAHEGMDQDKMPAPVHQIC----- 445
QY 298 KGSAGVTPPLAVTLDKGLQAPDNPALNTLLKQTLGKDTQHYLAHASSD--GSQHL- 354
DB 446 -----PAVCVLMK--LSFDEEH--RHAMNELGGLQIA 474
QY 355 -LLDNKGHLFDIKSTATSYSVLENSHPGEGIKGLAQAGTSVVDGSGKISLGSQTQSH 413
DB 475 ELLQVDCMEYGL--TNDHYSVTLRYAG-----MALTNLTFFGDVAN 513
QY 414 NKTMLSQPGEAHRSLLGIWCHPAGAAPQGESIRLHDDKIHILHPELVGWSADKDTHS 473
DB 514 KATLCSMKG--CMRALV-----AQLKSESEDLQOVIASVLIR--NLS--WRA---DVNS 557
QY 474 QLSRQADGKLYAL-----KDNRTLQ-----DNKSSEKLVNDKIKSYSV-- 513
DB 558 KKTIREVGSVKALMECALEKVEKSTLSVLSALNLSAHCATENKADICAVDGAFLVGT 617
QY 514 ----DQRGQVAILTDTGRRHMSIMPSLDASPESHISLSLHFADAHQCLL--HGKSELEAQ 568
DB 618 LTYRSQNTNLAIIES--GGGILRVSSLIATNEDHQRI--LRENNCLQTLLOHLK----- 668
QY 569 SVATSHGLVVDSEGLKF--SAAIPKQGDGNELKMKAMPOHALDEHGHQHOISGFPHD 626
DB 669 ----SHSLTIVSNACGTLWNLSARNPKD-----QELAWD--MGAVSMLNLIHS 711
```

```
QY 627 DHQQL---NALVKNNFRQQHACPLGNDHQPHQGNWLTDLALVIONQLGLHHTNPPEHEI-L 682
DB 712 KHRMIAMGSAALRNLMANRPARYKQDANIMSPGSSLP-----SLHVRKQKALEAEL 762
QY 683 DMGHLGSLAQEGKLYFDQLTGKWTGAESDCQL--KKGL--DGAAYLLKDGGEVKKRLNIN 739
DB 763 DAHLSET-----FDNIDMLSPKASHRSQRRHKQNDLYGYAFDANRHHDSRSNFN 813
QY 740 QSTSYKHGTENVFSLPHVRNKPPEPDALQGLNKKDDKAQAMAVIGNVKYLALTEKGDIRS 799
DB 814 TGNMTVLSPLYNTVLPSSSSSRGSLDSSRS--EKDRSLERERGIGLSAYHPTTIENACTSS 872
QY 800 FQKPGTQQLERPAQTLSREGISGELKDIIHVDHKQNYALTHEGEVPHQPREMONGAES 859
DB 873 ---KRGIIQITTTAAQIAK---VMEEVSAIHTSQDDRSSASTTE---PHCVADDRSAARRS 923
QY 860 SSWHKLK---LPQSESKLKLDM-----SHEHKPIATFED--GSQHLQKAGGWH 903
DB 924 SASHTSNTVNTFKSENSNRTCSMPYAKVYKRSNSDLSNSVTSSDGYGKRGQMKP--SVE 982
QY 904 AYAAPERGPLAVGTSGSTVFENL-----MQGVKGVIPGSGLTVLKLSAQ---TSGMTGA 955
DB 983 SYSEDDSEKFCYGYPADLAHKIHSANHMDNDGELDTPIYNSLYKSYDEQLNSGRQSPS 1042
QY 956 EGRKVS-----KFSERIRAYAFN---PTWSTPRPIKNAAYATQHCWQREGLKPL 1003
DB 1043 QNERWARPKHVIDEIKQNEQROARSQNTSYVYSENTDDKHLKFPQHFQOQ---ECVSP- 1099
QY 1004 YEMQGALIKQLD---AHNVYRNAPOPDLOSLETLDLGEHGAELNDKMRFRDELOSA 1059
DB 1100 YRSRGTSGETNRMGSSHAQNQNVNQLQBEDDYEDDKPTNYSEYSEEQHEEBERPT 1159
QY 1060 TRSVTVLQGHQGV-----LKSNGEINSEFKSPGKALVQSFNVNRSQD-----LSKS 1107
DB 1160 NYSIKYNEEKHVDQPIDYSLKYATDILSSQKPS-----FSFSKNSAQSTKEHLSPS 1213
QY 1108 LOQAVHATPSSASKIQLSMLGHFVSGVDMSHQK---EIPLGRQRDPNDKTALTCKRLI 1164
DB 1214 SENT--AVPPS--NAKRNQL--RPSSAQNRNGTQKGTTCVVP-----SINGETIQT--YCV 1262
QY 1165 LDT-VTIGELHELADKALVSDHKPDADQIKOLRQOFTDLREKREYESNPVKHYTDMGFT 1223
DB 1263 EDTPICFSRCSLSSLS--ADDEIGCDQTTQEADSANTLQTAEVKENDVTRSAEDPATE 1320
QY 1224 NKALEANYDAVKAFINA-----PKHEHGVNLTTRTVLESQGSAAELAKK-----LKNLT 1272
DB 1321 VPAVSQARAKPRLQASGLSSSTRINKAVESSGAKSPKSGAQTPKPPEHVQETP 1380
QY 1273 L-----SLDSGESMSFSRSY-----GGGVSTVFVPT----- 1298
DB 1381 LVFSRCTSVSLDSFSSRSIASSVQSEPCSMVSGIISPSDLPSQQTMPPSRSKTPPP 1440
QY 1299 -----LSKVPVPVPIPGAGITLDRAYNLSFRTSG---GLNVSGFRDGGVSGNIMVAT 1348
DB 1441 PPQTVQAKREVPKSKVPAE-----KRESGPKQTAVNAVOR-----VQVLP 1482
QY 1349 GHVMPVMTCKKTSAAG--NASDWLSAKHKISP-----DLRIGAAVSGTLOGTLQNSLKFK 1401
DB 1483 DVDTLHFATESPTDFGSCSSLSALSLEDEFFTKQVLELAIMPPVQ----- 1528
QY 1402 LTEDLPFGFIHGLTHGTLTPAELLQKGIHQMGKSKLTFS---VOTSANLDL----- 1451
DB 1529 --END-----NGNETESEQPEESNQDEVEKPDSEKOLLDDSDDDILEBECI 1577
QY 1452 -----PAGINLNEGSKPNGVTARVSAGLSASANLAAGSERSTTSGQFSGTTSAS 1502
DB 1578 ISAMPTKSSRKAKLAQATSKLPPPVARKPSQLPVYKLLPAQNRLQAKHVSF----- 1630
QY 1503 NNRPTFLNGVGAGANLTAALGVAHSSTHEGKPGVIGFPFAFTSTNVSAALADNRSTSQISL 1562
DB 1631 -----TPGDDVFRVYCVGETPI-----NFTSTATSLDLTIESPPN 1665
QY 1563 ELKRAEPVTSN-----DISELTSTLKGKFKDSATTKMLAAL--KELDDAKPAEQLHILOQ 1615
```

Db 1666 ELATGSDVRAGIQSGEFERDITPTGRSTDDAQRGKISSIVTPDLDDNK-ABEGDILAE 1724
QY 1616 HFSKADVGVDERVEARNLKLVIROQAADSHS-----MELGSASHSTT 1659
Db 1725 CINSAMPKG-KSHKPPR-VKKIMDQVQQASSTSSGANKNQVDTKKKKPTSPVKPMPQNT 1782
QY 1660 YNNLSRINDGIV-----ELLKHFDALPA--SSAKRLGEMM-NNDPALKDII----- 1705
Db 1783 YRTRVRKNTDSKVNVTETFSNDKDSKKPSLQTNAKAFNKLNNEDRVRGTFALDSPH 1842
QY 1706 --KQLOSTFP-----SSASVSMELKDLRQTEKALIDGKVGREEVCLFQDRNNLR 1755
Db 1843 HVTPIRGTPCYFSRNDLSLSDDDDDVLSREKAE--LRKGKESKDS-----EAKVTCR 1895
QY 1756 VKSVSVQSVSKSE-GFNTPALLLGTSNAAAMSMERNIGTFNKYGQD----- 1802
Db 1896 PEPNSSQQASKSQASIKHPA---NRAQSKPVLOKQPTFPQSSKQDPRGAATDEKLNQL 1952
QY 1803 --QNTPRRTLEGGIAQAANPQVASALTDLKK 1832
Db 1953 AIENTVCFSRNSSL-----SSLSDIDQE 1976

RESULT 13

BPEA HUMAN

ID BPEA HUMAN STANDARD; PRT; 5171 AA.
AC Q94833; Q8N1T8; Q8N8J3; Q8WK9; Q96AK9; Q96DQ5; Q9H555;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bullous pemphigoid antigen 1, isoforms 6/9/10 (Trabeculin-beta)
DE (Bullous pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein)
DE (Dystonia musculorum protein)
GN BPAG1 OR DMH OR DT OR KIAA0728.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORM 6), AND TISSUE SPECIFICITY.
RC TISSUE=Keratinocytes;
RX MEDLINE=21839111; PubMed=11751855;
RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;
RT "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1) including the domain structure closely related to MACF (microtubule actin cross-linking factor).";
RT J. Biol. Chem. 277:6682-6687(2002).
RN [2]

RP SEQUENCE OF 1342-5171 FROM N.A. (ISOFORM 10).

RC TISSUE=Ductal epithelium;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg K., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ushin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]

RP SEQUENCE OF 1702-4156 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain, Placenta, and Tongue;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Isogai T., Nishi T., Ota T.,
RA Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Kikuchi H., Masuho Y.,
RA Nagai K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 3535-5171 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual curation of 330 KIAA cDNA clones.";
RT DNA Res. 9:99-106(2002).
RN [5]
RP SEQUENCE OF 4107-5171 FROM N.A. (ISOFORM 6).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
RT DNA Res. 5:277-286(1998).
RN [6]
RP SEQUENCE OF 4031-5171 FROM N.A. (ISOFORM 9).
RC TISSUE=Brain;
RX Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP DISEASE.
RC TISSUE=Keratinocytes;
RX MEDLINE=92011493; PubMed=1717441;
RA Sawamura D., Li K., Chu M.-L., Uitto J.;
RT "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences deduced from cloned cDNAs predict biologically important peptide segments and protein domains.";
RT J. Biol. Chem. 266:17784-17790(1991).
CC -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing intermediate filaments to the inner plaque of hemidesmosomes. The proteins may self-aggregate to form filaments or a two-dimensional mesh (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=10;
CC Name=6; Synonyms=EA;
CC IsoId=Q94833-2; Sequence=Displayed;
CC Name=1;
CC IsoId=Q03001-1; Sequence=External;
CC Name=2;
CC IsoId=Q03001-2; Sequence=External;
CC Name=3; Synonyms=1e;
CC IsoId=Q03001-3; Sequence=External;
CC Name=4;
CC IsoId=Q03001-4; Sequence=External;
CC Name=5;
CC IsoId=Q03001-5; Sequence=External;
CC Name=7; Synonyms=EB;
CC IsoId=Q8WXXK8-2; Sequence=External;
CC Name=8;
CC IsoId=Q03001-6; Sequence=External;
CC Name=9;
CC IsoId=Q94833-3; Sequence=VSP_005068, VSP_005069;
CC Name=10;
CC IsoId=Q94833-1; Sequence=VSP_005066, VSP_005067;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and

```

CC cultured keratinocytes.
CC -!- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid
CC [MIM:600088], an autoimmune subepithelial skin blistering disease.
CC -!- SIMILARITY: Belongs to the plakins or cytokeratin family.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- SIMILARITY: Contains 29 spectrin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF400226; AAL62061.1; -.
CC EMBL; BC016991; AAL6991.1; -.
CC EMBL; AK055189; BAB70870.1; ALT INIT.
CC EMBL; AK094883; BAC04449.1; ALT INIT.
CC EMBL; AK096713; BAC04848.1; ALT INIT.
CC EMBL; AB018271; BAA34448.2; -.
CC EMBL; ALJ37008; CAC12899.1; -.
CC HSPF; P02631; IRRO.
CC Genew; HGNC:1090; BPAG1.
CC MIM; 113810; -.
CC MIM; 600088; -.
CC GO; GO:0005737; C:cytoplasm; ISS.
CC GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
CC GO; GO:0045104; P:intermediate filament cytoskeleton organiza. .; ISS.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR003108; GAS2.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR02017; Spectrin.
CC Pfam; PF00036; efhand; 2.
CC Pfam; PF02187; GAS2; 1.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00435; spectrin; 29.
CC ProDom; PD000012; EF-hand; 1.
CC SMART; SM00054; EFh; 2.
CC SMART; SM00243; GAS2; 1.
CC PROSITE; PS00018; EF_HAND; 2.
KW Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Structural protein; Cytoskeleton; Cell adhesion; Calcium;
KW Calcium-binding; Alternative splicing.
FT REPEAT 264 341 SPECTRIN 1.
FT REPEAT 349 444 SPECTRIN 2.
FT REPEAT 450 550 SPECTRIN 3.
FT DOMAIN 564 616 SH3.
FT REPEAT 935 1031 SPECTRIN 4.
FT REPEAT 1481 1581 SPECTRIN 5.
FT REPEAT 1715 1814 SPECTRIN 6.
FT REPEAT 1930 2008 SPECTRIN 7.
FT REPEAT 2071 2174 SPECTRIN 8.
FT REPEAT 2184 2282 SPECTRIN 9.
FT REPEAT 2294 2382 SPECTRIN 10.
FT REPEAT 2404 2502 SPECTRIN 11.
FT REPEAT 2513 2611 SPECTRIN 12.
FT REPEAT 2839 2940 SPECTRIN 13.
FT REPEAT 2950 3047 SPECTRIN 14.
FT REPEAT 3059 3156 SPECTRIN 15.
FT REPEAT 3168 3268 SPECTRIN 16.
FT REPEAT 3277 3376 SPECTRIN 17.
FT REPEAT 3386 3484 SPECTRIN 18.
FT REPEAT 3495 3594 SPECTRIN 19.
FT REPEAT 3716 3816 SPECTRIN 20.
FT REPEAT 3825 3925 SPECTRIN 21.
FT REPEAT 3935 4035 SPECTRIN 22.
FT REPEAT 4044 4142 SPECTRIN 23.
FT REPEAT 4153 4250 SPECTRIN 24.
FT REPEAT 4262 4362 SPECTRIN 25.
FT REPEAT 4372 4470 SPECTRIN 26.
FT REPEAT 4480 4578 SPECTRIN 27.

```

```

FT REPEAT 4596 4684 SPECTRIN 28.
FT REPEAT 4732 4790 SPECTRIN 29.
FT CA_BIND 4798 4810 EF_HAND 1 (POTENTIAL).
FT CA_BIND 4834 4846 EF_HAND 2 (POTENTIAL).
FT DOMAIN 126 160 COILED COIL (POTENTIAL).
FT DOMAIN 312 377 COILED COIL (POTENTIAL).
FT DOMAIN 402 447 COILED COIL (POTENTIAL).

Query Match 2.1%; Score 202; DB 1; Length 5171;
Best Local Similarity 18.1%; Pred. No. 0.53;
Matches 364; Conservative 310; Mismatches 720; Indels 616; Gaps 94;

Qy 143 DMGRPMVKGSGEDKVPYTOQ---KRHLNFMGQMRQTMLSKMAHPASANAGDRLQHSPP 199
Db 1138 EKAGKPPF---SKQKISSEISTKKEQLSEALQTIQLFLAKHGDKMTDEERNELEKQVK 1193

Qy 200 HIPGSHHEIKEEPPVGSTSKATTAHADRAVEITAEQDDDDSEFQOLHQOR-----LAREEN- 252
Db 1194 TLOESYNLFFSESLKQLQESQTSGDVKE---EKIVAEROQOEYKBEKLOGICDILLTQFENR 1250

Qy 253 ---PPOPPKLGVATPISARFQPK-----LTVAESVLEGTDTTQSPKPKQSMKLGSGA 302
Db 1251 LIGHQEAFMIGDGTVELKVKYQSQEELQKMQGSAQALAEVVKNTENFLK-ENGEKLSQE 1309

Qy 303 GVTPLAVTLDKGLQL-----APDNPALNTLLKOTLGKDTQHYLAHASSDGSQHL-- 354
Db 1310 DKALIEQKNEAKIKCEQLNKAQSKKELDKVVTVAIKEETKEKVAAVKQLEESKTKIEN 1369

Qy 355 LLDNKGHLFDIKSTATSYSLVNSHPFEEKGLAQAGTGSVSDGKSGKSLGSG----- 408
Db 1370 LLD---WLSNVDKDSERAGTKH-----KQVTEQNGTHFQEGDGKS---AIGSEDEVNG 1416

Qy 409 -----GTQSHNKTMLSOPGEA-HRSLTLGIWHPAGAARPOGESIRLHDDKIHI 456
Db 1417 NLLETVDVGQVGTQENLNQYQVKVKAHEKITS---OHOAVIITASAQVLLKQGOYL 1473

Qy 457 -----LHPGLVMQSDAKD--THS-----OLSRQ- 478
Db 1474 SPEBEKELQNMKELKVHYETALAESEKMKMLTHSLOEELEKFPDADVTEPEHMLQSEQE 1533

Qy 479 -----ADGKLYALKDNRTL-QNLSNKK-----SSEKLVDKIKSYSDVQGOV 519
Db 1534 LENLEAGADDINGIMTKLKEQKSFSEDIVSHKGLDLYITISGNRVLEAAKSCSRDGGKV 1593

Qy 520 ALLTDTPTGRHKMSIMPSLDASPSHISL-----SLHFADAHQGLLHGK 562
Db 1594 ----DTSATHR-EVQRKLDHATDFRSLSYKCNVGNLXLDVLDKYQHYEDASGGLAGL 1648

Qy 563 SELEA-----QSVATSHGRLVVADSEKGLFSAAIKQGDGNELKMKAMPQHALDBHFG 615
Db 1649 QACEATASKHLSPIAVDPKNLQRLQLEETKALQGISSQVAVVE-KLKKTAEVLLDA--- 1704

Qy 616 HDHQISGFFHDDHGQLNALVKNPNFRQOHACPLGNDHQHFHGWNLTDALVDNQLG---L 671
Db 1705 -----RGSL-----LPAXNDIQ-----KTLDIDVGRYEDL 1729

Qy 672 HHTNPPEHILDMGHLSLALQEGKLHYFDLTQKGTG-AESDCK-OLKGLGDAAVLLK 729
Db 1730 SKSVNERNEKQITLIRSLSVQDGLDEMLD-----WNGVNESSLKEQGVPLNSTA--LQ 1782

Qy 730 DGEVKRLINQOS-----TSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGV 785
Db 1783 DIISKIMLEQDIAGROSSINAMNEKVKFMEITD-PSTASSLQAKMKDLARPSEASHK 1841

Qy 786 NK-YLAUTE--KGDIRSFOIKPGTQO--LERPAQTLRSREGISGELKDI----- 828
Db 1842 HKETLAKMEELKTKVELFENLSEKLTQFLTKTQALTEVDVPG--KDVTELSTQYMQBSTS 1899

Qy 829 -HYDHNKONLYALTHEGVRPHQ-PREAWQNGAESSSWHKLALPOSES---KLKSL-DMSHE 882
Db 1900 EFIEHKKHL-----EVLHSLKETSISHLGSPDK--ALVLEKINNLSKKFKEMEDTIKE 1950

Qy 883 HKPIATFEDSGSHQLKAGGWHAYAAPERGLAVGTSGTSQTVFNRLMQGVKG-----KVI 936

```


Query Match 2.1%; Score 199.5; DB 1; Length 3590;
Best Local Similarity 19.3%; Pred. No. 0.39;
Matches 426; Conservative 268; Mismatches 793; Indels 725; Gaps 114;

QY 5 SLGTEHKAHVHTAAHPNPGVGHVALOOGSSSPQWAAASLAAEGKRGKMPRIHOPSTAA 64
DB 1379 SLDIKKGAQVTVAGRYAEHGEVSTQGYTYSAD--AIALAAQVTFQG-----GA 1426
QY 65 DGISAAHQKKSFSLRGCLGTKKFSRSAPQGPQGTTH--SKGATLLDLLARDDDGETOHEAA 123
DB 1427 ANLTSRHDTRSNKIR-LMGPLQVNVAGGAVSNTGNLKVREGVTV--TAAFPDNETGAEMV 1483
QY 124 APDAARLTRSGVKRRNMDMAGRMVK-----GGSG 155
DB 1484 A-KSATLLTTSAAARN-----AGKMQVKEAATIIVAASVSNPGTFTAGKDIITVSRGGFDN 1536
QY 156 EDKVPYTOQ-----KRHOLNN-----FCQMRQTMLSKMAHPASANAGDRLQHS 197
DB 1537 EGMESNKDIIKTEQFNGRVLDAKHDLTVTASQADNRSLKAGHDFTVOA--QRIDNS 1595
QY 198 PPHIPGSHHEIKBEPVSGTSKATTAHADRVEIAQEDDDSEFQQLHQOQLARERENPPOPP 257
DB 1596 GTMAAGHDATLKAHLRNTGQVWACH-----DIHI INSAKLENTGRVDA 1639
QY 258 KLGVA TPISARFPQKLTAVAE-----SVLEGT-----TTQSPLKP--QSMILKSGGAGVTP 306
DB 1640 RNDIALDV-ADFTNTGSLYASHDATLTLAQGTQRDVVDQHLPLVABEGTLRVKAKSLTT 1698
QY 307 LAVTLDKGLK---QLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKGHL 362
DB 1699 ELETGNPGLTAEVOENTDNQAI-----VVGKDLTLSSA-HGNVANEANALLWAAGEL 1751
QY 363 FDIKSTATSYSLVHNSHPEKIGKLAQAG-----TGSVSVGKSGKISLG-----SGTQS 412
DB 1752 -----TVKAQNI TNK-----RAALIEAGGNARLTAAVALLINKLRIPAGEDMHLDPRI 1800
QY 413 HNKTMLSOPGGAHRS-----LLTGIWOHPAGAARP-----442
DB 1801 ENTAKLS--GEVQRKGVQDVGGBHGRWSGTGYVNYWLRAGNKGKAGTIAAPWYGGDLTA 1858
QY 443 -----QGESIRLH-----DKIHLHPGLVQWQSAKDTHSLSRQADGK-----482
DB 1859 EQSLIEVKOLYLNAGARKDEHRHLNE--GVIQAGG---HGHIGDNDNRSVVTVTSAM 1913
QY 483 -----LYALKDNRTLQNLSDN-KSSKLVKDKIKSYSDVQGVVAIITDTPGRHKM 531
DB 1914 BYFKTLPVSLTALDNRAGLSPAIFWQSTYELLD-----YLLDQNRVEYIWLGYTYTEW 1969
QY 532 SI-----MPSLDASPESHISLSLHFADAHQ-----LLHG---KS 563
DB 1970 SVNTLKNLDLGYQAKPAPTAPMPKAPF--LDLRGHTLESAGRKIFGEYKKLQGEYEKA 2027
QY 564 ELEAQSV-AISHGLVNVADSGKLFSAAI-----PKQGDGNELMKMKAMPQHALDEHFGH 616
DB 2028 KMAVQAVEAYGEATRRVHQDQGRYKALGMDAETKEVDG-----IIOQFAA 2075
QY 617 DHOISGFFHDDHGOLNA---LVKNNFRQO--HACPLGNDHOPHGWNLTDALVIDNQLG-- 670
DB 2076 DLRTVYAKQADQATIDAETDKVAQYKSIQIDAVRL---QAQPG-RVTLAKALSAALGAD 2131
QY 671 ---LHHTN-----PEPHEILDMGHLSLALQEGKLYFDQLTKGM 707
DB 2132 WRALGHSQLMQRMQDFKAGKRGAEIATFYPKEQTVLAAG--AGLTLSNGAIH-----NGE 2183
QY 708 TGAESDCKQ--LKKG-----LDGAAYLLKD-GEVKRLINQSTSSIKHGTENTVPSLPHV 758
DB 2184 NAAQNRGPEGLKTAHSATSVSGSFDALRDVGLKLEKLDIDDALAA-----VLVNFHI 2236
QY 759 -----RNKPE-----PGDALQGLKNDKQAKAMAVI 783
DB 2237 FTRIGAACTSLADGAAGPALARQARQAPETGMDVARDGLGSADALASLASLDAAGLEVS 2296
QY 784 GVNKYALTEKGDIRSFOIKPFGTQOLERP---AQTLRGEGISBELKD---IHVDHKQNLV 837

DB 2297 G-RRNAQVADAG-----LAGPSA VAPAAVGAADVGVEPVTGQVQDQVVAVAGLEQFVA 2348
QY 838 ALTHEGEVPHOPREAWQNGAESHHKIALPQSSSKLSLSDMSHEHKFIATFEDSQSOL 897
DB 2349 TVRVAPPAVALPR-----PLFETRIKFIQDOKFY-----GSRYPF 2383
QY 898 KAGWHAAYAAPERGLAVGTSGSOTVF-----NRLMOQVKGV-IPGSGLTVKLSAQTTGG 951
DB 2384 EQIGY-----KPDRAARVAGDNYFTTLVREQVRALGGYESRLPVRGVALVAKLMDSAGT 2439
QY 952 MTGAEGRKV-----SSKFSERIRAYAFNPTMTPTPIKNAAYATQHWGREGKLPY 1004
DB 2440 VGKALGLKGVAPTAAQQLKQADRFVWYVDVVIDQKVLAPRLYLTE---ATROGITITQY 2496
QY 1005 EMOCALIK-----QLDAHNVRHNAPOPDLQSKLETLDLGHEGAELLNDMKRFRDELEQ 1057
DB 2497 AGGALLIAGSGDVTNTDGHV--SSVNGLIQGRSVKVDAGK-GKVYVADSKGAGGIE- 2552
QY 1058 SATRSVTVLGQHQGVLSNGEINSEFPKPGKALVQSPNVRNSGDLKSLQQA VHA TTP 1117
DB 2553 -ADDEVDSGRDIGI--EGGKLR-----GKQVPLKADTVKVATSM 2589
QY 1118 SAESKLOSMUGHFVSAGVD--MSHQGEIPLGRQDPNDKTALT--KSLILDTVTIGELH 1174
DB 2590 RYDDK-----GRLAARGDGDALDAQGQLHIEAKRLETAGATLKGKGVKLDVDDVKLG VY 2644
QY 1175 ELADKAKLVSDHKPDADQIKOLRQOFDTLREKRVESNPVHYTDMGFTH--NKALEANY 1231
DB 2645 EAGSS-----YEN---KSTPLGSLFALLSSTETETNQ 2673
QY 1232 DAVKAFINAFKKEHGVNLTTRTV-----LESQGSALAKLAKLNTLILSLDSGESMSFSR 1285
DB 2674 SA-----HANHYGTIEAGTLEGKMNLEIEGGSVDA---AHTDLSVARDARFKAA 2722
QY 1286 SYGGVSTVFPVPTLSKKVPVPIPGAGITLDRAYNLSPRST--GGLNVSVFGRDGVSN 1344
DB 2723 DFAHAHEKDVQRLSLGAKV---GAG---GYEAGFSLGSESGLEHAHAGR--GMTAGA 2771
QY 1345 MVATGHVMPYMT-----GKTSAGNASDMLSAKHKISPDIL---RIGAAV 1386
DB 2772 EVKGYRASHEQSSSETEKSYRNANLNFGGGVEAGNVLDIGGA-----DINRRYGGAA 2825
QY 1387 SGTLLQTLQNSLKFELTEDELPFGFIHGLT-----HGTLPAPALLQKGI 1429
DB 2826 KGN-AGT-EAALRMRAKKVESTKVVSEQTSSQSCSVSEVASTASARSLTLTAATFELGDSV 2883
QY 1430 EHQMKQSGKL-----TFSVDTSAMLDLRAGINLNEGDSKPNQVTA-----1469
DB 2884 AQNVEDGREIRGELMAAQA AAEATQLVTADTAA--VALSAGISADFSHRSRSTQNTQYL 2942
QY 1470 -----RVSAGLSASANLAAGSR--ERSITTSQGFSTTSASN 1503
DB 2943 GGNLSIATEGDTATLVGAKFGGGQVSLKKAASVNLMAAESTFESYSESHNFHASADANL 3002
QY 1504 NRPTFLNGVAGANLTAALGVVAHSSTHE-GKPVGIFPAFTSTNVSA--LALDNRTSQSIS 1561
DB 3003 GANAVQAGVGLG--LTAGMGTSHQITNETGK-----TYAGTSVDAANVSID--AGKDLN 3052
QY 1562 LELKRAEPVTSNDISELTSTLGHKHFQDSATTKMLAALKELDADKAPQLHLQOHHFSAKD 1621
DB 3053 LSGSRVR-----GKHVLDVEGDINATSKQ-----3077
QY 1622 VUGDER-YE-----AVRNKLKLVIRQOAAASHMELGSA--SHSTTTNNLSRI 1666
DB 3078 ---DERNYSNGGWDASAGVAIQN-RTLV-----APVGSAGFNFTNTEHDN--SRL 3122
QY 1667 NNDGIVELL-----HKHFDAALPASSAKRGLGEMMNDPALKDI IKQLQSTPSSASVS 1719
DB 3123 TNDGAAGVADSGLTGHVKGDNLTGATLADLSKGN---LK-----VDGAVNA 3168
QY 1720 MELKQGLREQTEKAILDKGVREEVGLFQDRNNLRVKSVSQSVSKSEGFNTPALLLG 1779

Db 3169 QNLKD-YRDK-----DGGSGGLNVG-----ISSTTLAPTGVAFG-----RVAG 3206

Qy 1780 TSNRAMSMERNIGTINFKYQDONTFRBFTLEGGIAQANPOVASALTDLKK 1831

Db 3207 EDYQAEQRATIDVG-----QTKDPARLVGGGVKGTINQDAAQATVVQR 3250

RESULT 15

CING_MOUSE STANDARD; PRT; 1191 AA.

AC P59242;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cingulin.

GN CGN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.P., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.

RT Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- FUNCTION: Probably plays a role in the formation and regulation of the tight junction (nJ) paracellular permeability barrier (By similarity).

CC -1- SUBUNIT: Homodimer. Interacts with TJP1/ZO-1 (By similarity).

CC -1- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but does not abolish colocalization with ZO-1 (By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; BC042459; AAH42459.1; ALT_INIT.

DR MGD; MGI:1927237; Cgn.

DR GO; GO:0005923; C:tight junction; NAS.

DR GO; GO:0003779; F:actin binding; ISS.

DR InterPro; IPR002928; Myosin tail.

DR Pfam; PF01576; Myosin tail_1.

DR Tight junction; Coiled coil.

KW DONAIN 1 345 HEAD.

FT DONAIN 343 1150 COILED COIL (POTENTIAL).

FT DONAIN 1149 1191 TAIL.

FT DONAIN 357 830 GLU-RICH.

FT SITE 42 56 ZIM.

SQ SEQUENCE 1191 AA; 136446 MW; 43EA20F49AF5E7D9 CRC64;

Query Match

2.1%; Score 199; DB 1; Length 1191;

Best Local Similarity 19.4%; Pred. No. 0.073;
Matches 265; Conservative 190; Mismatches 491; Indels 422; Gaps 66;

Qy 16 TAAHNPVGH--GVALQGSSSS-----SPQNAAS--LAAEKNRGMKPRIH-QSTAA 64

Db 113 TSDEEPKHNGKLIRSQASLTGLAFMSPNRSTLSLELAPKPTSSINTIDAPLSSV 172

Qy 65 DGISAAHQOKKPSLRGCLGTKKFSRSPAQGGQGTTHSKGATLRDLARDGGTQHEAAA 124

Db 173 DSLINKFDSQKGGQVGRGTGR--TRTLPEQ---RKRSQSLDSRLPRD---TREERH 223

Qy 125 PDAARLTRSGVKRRNMDMAGRPVMVKGSGGDKVPTQOKRHQNLNFGQWRQM----- 178

Db 224 QSANHWTR--GTKYDNHVDSSKNP-----SQKSPFSFSRSRQTDQWVLOS 268

Qy 179 LSKMAHPASAN-----AGDRLQHSPPHPIGSHHEIKKEPVGSGTSKATTAHADRVETIAQ 231

Db 269 FEETROPAMVQFKSTPDLRDQRETAP---PGADHVKATYIGI-----LRE 312

Qy 232 EDDDSFQOLHQOQLARERENPPQPKLGVATPISARF---QPKLTAVAESVLEGTDTTQ 288

Db 313 GSSESEASVRRKVSLLVEQMOP-----LGMVSPASTKALAGQAELTRKMEELQKLD--- 364

Qy 289 SPLKQSMKLGSGAGVT-----PLAVTLD--KGKQLAPNPPALNTLLKQTLG 335

Db 365 EEVKRQKLEPSRVGLERQLEKAEECHRLQELLERRKGEVQSSKELQNMKLLQGBEG 424

Qy 336 -----KDOHYLAHASSD--GSOHL---LLDNK---GHLFDIKSTATSYSLVHNSH 379

Db 425 LRHGLBAQVKELQKHKQSPDSGKESLKLDDLTRELLLELEGKQVVEQLRERERE 484

Qy 380 PGEIKGLKAQAGTG-----SVSDVGSKSKISLGSQTQSH-----NKTMLSOPGE 423

Db 485 LTALKGALKKEEVASHQVEHVLQYQRTDTEQLRRSQMDATODHAALAEAKMXSLVRE 544

Qy 424 AHRSL-----LTGIWQ-----HPAGAAPQGESIRLH-----DDKIHLHPELGV 463

Db 545 LQRELEBETSBETHGWSQMPQKNKEELRATKQELLQRLMEKEEBELGEMVELQRLDLEQ 604

Qy 464 WQSAADKTHQSLSRQADGKLYALKDNRTLQNLSDNKSSEKLVKDKIKSYSDVQGGQVAILT 523

Db 605 ARASTRDTH-----QVEELK-----KELRRTQGLKELQAEQQNQ----- 639

Qy 524 DTPGRHKMSIMPSLDASPESHISLSHFADAHQGLHKGSELBAQSVASHG--RLVVADS 582

Db 640 EVTGRHNRQVL-----EKQALAREAD-----RGR-ELEQNLQQLKTLQQLRQDC 685

Qy 583 EGKLFSAATPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHCOLNALVKNFRQ 642

Db 686 E-----EASKAKVASETEAMML-----GQRATVETTTURET 716

Qy 643 HACPLGNDHGFPGWNLTDALVIDNQLGHLHTNPPEHEILDGMHGLSLAQEGKLHYFDQ 702

Db 717 QE-----ENDEFR-----RRILGLEQQLKEARGLAEGGEAVEARLD--KVHRL-- 758

Qy 703 LTKGWTGAESDCQKXKGLDGAAYLLKDGVEKRLINQSTSSIKHGTFENVPSLPHVRNKP 762

Db 759 -----EVEKQQLBEALNAQOE--EENGL-----AAAKRALE----- 787

Qy 763 EPGDALQGLNKDDKAQAMAVIGNVKYL---ALTEKGDIRSFOIKPGTQQLERPAQTLSR- 818

Db 788 -----VLDQAQGLARLQGOEQQALNPALEBEGKQRE--ALRRSKAELEEQKRLNRT 838

Qy 819 -EGISGELKDIHVDHKQNLQYALTHEGEVFPHPREAWONGAESSSWHKIALPQSESKLSL 877

Db 839 VDRINKELEIQIDD-----SKLALQQLAQAME-- 865

Qy 878 DMSHEHKPIATFEDGSGHQHAKGWHYAYAPERGPVAVGTSGSQTFVFNRLMQGVKGVIP 937

Db 866 ----DYKEKAR--KEVADAQQAQKADWASEAEKNSG-----GLSRLQDELQRLQAL----- 910

Qy 938 GSGLTVKLSAQTGWTCAEGRKVSSKFSERIRAYAFNPTMTSTPRPIKNAAYATQHGQGR 997


```

Db 911 -----QTSQABRTARLDKELLAQRLOGLE-----QEA 938
Qy 998 EGLKPLYEMOGALIKOLDAHNVHRNAPQDIOQSKLET-LDLGERGAELLND-MKRFREDEL 1055
Db 939 ENKKRFODDKARQLKSLE-----EKVSRLEAELDEEKNTVELLTDVNRGRDQV 987
Qy 1056 EQSATRSVTVLGQHGVKNGEINSEBPKSPGKALVQSFNVNRSQDLSKSLQOAVHAT 1115
Db 988 DQLRTELMQERSARQDL-----ECDKISLERQNKDLKTRIASSEGFQ 1029
Qy 1116 PPSAE-SKLOSMLGHFYSAGVDMSHOKGEIPLGRORPNDKTALTCSRLLDVTITIGELH 1174
Db 1030 KPSASLSQLES-----QNLQERLQABEREKTVLQSTNRKLER-RVKELS 1074
Qy 1175 -ELADKAKLVSDHKPDAD-QIKLRQOQDITLRE--KRYESNPVKHYTDMGFTH--NKALE 1228
Db 1075 IQIDDERQHVNQDKQLTLRVKALKRQVDAAEEIEIRLDSLRKKAQRELEBEQHEVNEQLQ 1134
Qy 1229 ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAAELAKKLNLTLLSLD 1276
Db 1135 ARIKSLEK--DAWRK-----ASRSAAESA--LKQEGLSDD 1165

```

Search completed: July 7, 2004, 15:18:37
 Job time : 44.0314 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:13:22 ; Search time 81.8129 Seconds
(without alignments)
7088.403 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTLKKEGLEMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 O54581	O54581 erwinia amy
2	5529.5	58.5	1829	2 Q9KH44	Q9KH44 pantoea agg
3	5502.5	58.2	1835	2 Q9FCY7	Q9FCY7 erwinia ste
4	1831	19.4	1795	16 Q887C9	Q887C9 pseudomonas
5	1824	19.3	1795	2 O66101	O66101 pseudomonas
6	1198.5	12.7	1160	2 Q8RNV8	Q8RNV8 pseudomonas
7	941	10.0	1441	2 Q8VPK4	Q8VPK4 pseudomonas
8	325	3.4	1957	16 Q888R1	Q888R1 pseudomonas
9	302.5	3.2	1702	16 Q8XQB6	Q8XQB6 ralonstia s
10	283.5	3.0	9904	16 Q8NMQ6	Q8NMQ6 staphylococ
11	283	3.0	57	2 Q9LAW1	Q9LAW1 erwinia amy
12	275.5	2.9	6713	16 Q9U54	Q9U54 staphylococ
13	275	2.9	2514	16 Q9J30	Q9J30 neisseria m
14	273.5	2.9	6713	16 Q9J30	Q9J30 staphylococ
15	261.5	2.8	1877	4 Q8NEY1	Q8NEY1 homo sapien
16	258.5	2.7	9439	16 Q8CP76	Q8CP76 staphylococ

ALIGNMENTS

RESULT 1

ID	O54581	PRELIMINARY;	PRT;	1838	AA.
AC	O54581;				
DT	01-JUN-1998	(TrEMBLrel. 06, Created)			
DT	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)			
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)			
DE	DSPE.				
GN	DSPE OR DSPA.				
OS	Erwinia amylovora.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Erwinia.				
OX	NCBI_FaxID=552;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EA321;				
RC	MEDLINE=98115919; PubMed=9448330;				
RA	Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,				
RA	Conlin A.K., Collier A., Beer S.V.;				
RT	"Homology and functional similarity of an hrp-linked pathogenicity				
RT	locus, dspeP, of Erwinia amylovora and the avirulence locus avrE of				
RT	Pseudomonas syringae pathovar tomato."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330 (1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CFBP1430;				
RC	MEDLINE=98086111; PubMed=9426142;				
RA	Bogdanove A.J., Kim J.F., Wei Z.-M., Kolchinsky P., Beer S.V.;				
RA	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CFBP1430;				
RC	MEDLINE=98086111; PubMed=9426142;				
RA	Gaudriault S., Malandrin L., Paulin J.F., Barny M.A.;				
RT	"DspA, an essential pathogenicity factor of Erwinia amylovora showing				
RT	homology with AvrE of Pseudomonas syringae, is secreted via the Hrp				
RT	secretion pathway in a DspB-dependent way."				
RL	Mol. Microbiol. 26:1057-1069 (1997).				
RN	[4]				
RP	SEQUENCE OF 1-238 FROM N.A.				

Q9P9U6 xylella fas
Q8CH77 mus musculus
Q840G7 pseudomonas
Q9A0K5 streptococ
Q8IR22 drosophila
Q9PBE8 xylella fas
Q8WXQ6 homo sapien
Q9K0T0 neisseria m
Q9GJP6 homo sapien
Q8WY20 homo sapien
Q8G9X9 escherichia
Q8IVL2 homo sapien
Q8NUJ3 staphylococ
Q8ZDR6 versinia pe
Q8DOR8 versinia pe
Q8OI7 mus musculus
Q76891 drosophila
Q9IXS5 pasteurella
Q8VQ99 staphylococ
Q8KT64 photorhabdu
Q9CP79 homo sapien
Q9ZHL3 haemophilus
Q7VIE8 haemophilus
Q97205 leishmania
Q99GY4 staphylococ
Q9AHF9 fusobacteri
Q8XQZ5 ralonstia s
Q8CNU9 staphylococ
Q8EWS3 shewanella

RC	STRAIN=EA321, and ATCC 49947;	
RX	MEDLINE=98422475; PubMed=9748455;	
RA	Kim J.F., Beer S.V.;	
RT	"HrpW of <i>Erwinia amylovora</i> , a new harpin that contains a domain	
RT	homologous to pectate lyases of a distinct class.";	
RL	J. Bacteriol. 180:5203-5210(1998).	
DR	EMBL; U97504; AAC04850.1; -	
DR	EMBL; Y13831; CAA74156.1; -	
DR	EMBL; U94513; AAC62315.1; -	
DR	PIR; T18448; T18448.	
SQ	SEQUENCE 1838 AA; 198243 MW; 7D595F78130E8FD9 CRC64;	
Query Match 100.0%; Score 9448; DB 2; Length 1838;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSSPQNAASLAAGKNGKMPRIHQ	60
DB	1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSSPQNAASLAAGKNGKMPRIHQ	60
QY	61 STAADGISAHQOKKSFSLRGCLGTTKFSRPAQGGTTHSKGATRLDARDGCTQH	120
DB	61 STAADGISAHQOKKSFSLRGCLGTTKFSRPAQGGTTHSKGATRLDARDGCTQH	120
QY	121 EAAAPDAARLTRSGGVKRRNMDMAGRPVKGSGEDKVPQKRHQLNFFGQMRQTM	180
DB	121 EAAAPDAARLTRSGGVKRRNMDMAGRPVKGSGEDKVPQKRHQLNFFGQMRQTM	180
QY	181 KMAHPASANAGDRLOHSPPHIPGSHHEIKKEEPPVGSSTKATTAHADRVETAQEDDSEFQ	240
DB	181 KMAHPASANAGDRLOHSPPHIPGSHHEIKKEEPPVGSSTKATTAHADRVETAQEDDSEFQ	240
QY	241 LHQQLARENPPQPKLGVAFTPI SARFQPKLTAVAESVLEGTDTTQSPKFPQSM	300
DB	241 LHQQLARENPPQPKLGVAFTPI SARFQPKLTAVAESVLEGTDTTQSPKFPQSM	300
QY	301 GAGVTPPLATLDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHL	360
DB	301 GAGVTPPLATLDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHL	360
QY	361 HLFDKSTATSVLHNSHPGETKGLAQAGTGSVDGSGKI SLGSGTQSHNKTML	420
DB	361 HLFDKSTATSVLHNSHPGETKGLAQAGTGSVDGSGKI SLGSGTQSHNKTML	420
QY	421 PGEAHSLLTGIWHPAGAARPGESIRLHDDKIHLHPGLGVWQSAKDQTHS	480
DB	421 PGEAHSLLTGIWHPAGAARPGESIRLHDDKIHLHPGLGVWQSAKDQTHS	480
QY	481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQGVQVAILTDTGPRHKMS	540
DB	481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQGVQVAILTDTGPRHKMS	540
QY	541 PESHISILSHFADAHQGLLHGKSELEBAQSVAI SHGRLVVDSSGKLFSAAI	600
DB	541 PESHISILSHFADAHQGLLHGKSELEBAQSVAI SHGRLVVDSSGKLFSAAI	600
QY	601 KMKAMPQALDEHFGHDHQSIFPHDDHQLNALVKNFRQQAACPLGNDHPQ	660
DB	601 KMKAMPQALDEHFGHDHQSIFPHDDHQLNALVKNFRQQAACPLGNDHPQ	660
QY	661 DALVIDNQLGHTNPEPHEIILDMGHLGSLAQEGKLYHFDQLTGKWTGAES	720
DB	661 DALVIDNQLGHTNPEPHEIILDMGHLGSLAQEGKLYHFDQLTGKWTGAES	720
QY	721 LDGAAVLLKDGGEVKRLNIQSTSSIKHGTENVFSLPHVRNKPDPGALQ	780
DB	721 LDGAAVLLKDGGEVKRLNIQSTSSIKHGTENVFSLPHVRNKPDPGALQ	780
QY	781 AVIGVKNYALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKD	840
DB	781 AVIGVKNYALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKD	840
QY	841 HEDEVPHQPREAWQNGAESSSWHKLALPQESKLSLDMSSHEHKPTATF	900

DB	841 HEDEVPHQPREAWQNGAESSSWHKLALPQESKLSLDMSSHEHKPTATF	900
QY	901 GHVAYAAPBPGPLAVGTSGQTVFNRLMGVKGKVI PGSGLTIVKLSAQGTG	960
DB	901 GHVAYAAPBPGPLAVGTSGQTVFNRLMGVKGKVI PGSGLTIVKLSAQGTG	960
QY	961 SSKFSERIRAYAFNPTMTSTPRPIKNAAYATQHGQWQREGKLPLEYOG	1020
DB	961 SSKFSERIRAYAFNPTMTSTPRPIKNAAYATQHGQWQREGKLPLEYOG	1020
QY	1021 HNAPODLOSLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGQGV	1080
DB	1021 HNAPODLOSLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGQGV	1080
QY	1081 SEFKSPGKALVQSFNVRSGQDLKSLSQAAVHATPPSAESKLSQMLG	1140
DB	1081 SEFKSPGKALVQSFNVRSGQDLKSLSQAAVHATPPSAESKLSQMLG	1140
QY	1141 KGEIPLGRORDNDKALTAKSRLIILDTVTIGELHELADKAKLVSDHK	1200
DB	1141 KGEIPLGRORDNDKALTAKSRLIILDTVTIGELHELADKAKLVSDHK	1200
QY	1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGV	1260
DB	1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGV	1260
QY	1261 SAEALAKKNTLLSLDSESMSPSRSGYGGVSTVFVPTLSKKVPVPI	1320
DB	1261 SAEALAKKNTLLSLDSESMSPSRSGYGGVSTVFVPTLSKKVPVPI	1320
QY	1321 LFSRTSGGLNVFSGDGGVSGNIMVATGHDVMPYMTGKTSAGNAS	1380
DB	1321 LFSRTSGGLNVFSGDGGVSGNIMVATGHDVMPYMTGKTSAGNAS	1380
QY	1381 RIGAAVSGTLQGTQNLKFKLTDELPGFIHGLTHGTLTPAELLQK	1440
DB	1381 RIGAAVSGTLQGTQNLKFKLTDELPGFIHGLTHGTLTPAELLQK	1440
QY	1441 FSVDTSANLDRAGINLINEDGSKPNGVTARVSAGLSASANLAAG	1500
DB	1441 FSVDTSANLDRAGINLINEDGSKPNGVTARVSAGLSASANLAAG	1500
QY	1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIPPAFTSTNV	1560
DB	1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIPPAFTSTNV	1560
QY	1561 SLELKRAEPVTSNDISELTSTLGKHPKDSATTMKAALKELD	1620
DB	1561 SLELKRAEPVTSNDISELTSTLGKHPKDSATTMKAALKELD	1620
QY	1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGASHSTTYNNLS	1680
DB	1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGASHSTTYNNLS	1680
QY	1681 AALPASSAKRLGEMMNDPALXDI IKQLOSTPSSASVSMELKDG	1740
DB	1681 AALPASSAKRLGEMMNDPALXDI IKQLOSTPSSASVSMELKDG	1740
QY	1741 REEVGVLFODRNLNRKVSQSVSKSGENTFALLGTSNSAAMSMERN	1800
DB	1741 REEVGVLFODRNLNRKVSQSVSKSGENTFALLGTSNSAAMSMERN	1800
QY	1801 QDQNTFRRFTLEGGIAQANPQVASALTDLKKEGLEMS	1838
DB	1801 QDQNTFRRFTLEGGIAQANPQVASALTDLKKEGLEMS	1838

RESULT 2
Q9XH44
ID Q9KH44
AC Q9KH44;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)


```
QY 1656 HSTYNNISRNNDGIVELLHKKHFDALPASAKELGEMMNNDPKDIIKOLQSTPFS 1715
D 1653 HTTQYSLRLDENGIFILHHHFAALPPTSATRISNMANNPLKALIQKLSPTS 1712
QY 1716 ASVSMELDGLREOTEKAILDQKVGREEVGLFODRNLRVKSVSQSVKSEGFNTPA 1775
D 1713 ARVSMELDGLDQTEKAILDQKVGREELARFODRNLRVKSVSQSVKSEGFNTPT 1772
QY 1776 LLLGTSNAAMSMEINIGTINFKYQDQNTPRFTELEGIAQANQVASALTDLKEGLE 1835
D 1773 SILGASNSAGLSMGNGTINFKYQDQNTPRFTELEGIAQANQVASALSELKEGFE 1832
QY 1836 MKS 1838
D 1833 MKS 1835

RESULT 4
ID Q887C9 PRELIMINARY; PRT; 1795 AA.
AC Q887C9;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Avirulence protein Avre (Pto).
GN AVRE (Pto) OR PSPT01377.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016860; AA054899.1; -.
DR TIGR; PSPT01377; -.
SQ Complete proteome.
KW SEQUENCE 1795 AA; 195149 MW; 6C9BF0D71364A08C CRC64;

Query Match
Best Local Similarity 29.5%; Pred No. 2.8e-85;
Matches 564; Conservative 363; Mismatches 737; Indels 248; Gaps 75;

QY 55 PRIH-----QPSTAADGISAH-----QQKSFSLRGCLGT-----KKFSRS- 91
D 4 PSIHRTGSIITPTVTPDARAATDLQERAEQPRQRSSHL--SSVGRALKSVGKLFQSK 62
QY 92 APOQOPGT-----THSKATLDLLARDG-----ETQHEAAPPDARLTRSGV 136
D 63 APOQKAATPTTAKNVKTPPPASNVATPRN--KARESGFSNSPQNTHSAPKSI----- 113
QY 137 KRRNMDMAGRPVMYKGGSGEDKVPVTPQKRHOLLNFGQMTLSCMAHPASANAGRLQH 196
D 114 -----LRNHPNQASSG-----AQTHEIPEAPRKNLRVRFDLQ-----DRUER 154
QY 197 SPHPHPSHHEIKEEPVGSTSKATTADRVETIAQEDDDSEFQQLHOORLARENPPOP 256
D 155 SPSYLDSDNPMTDEAVANATRF-----RSPDSLQSGDGTFRISMLATDPDP 203
QY 257 ----PKLQVAT--PISARQPKLTAVAESVLGTDTTQSLPKPQMLKSGAGVTPL-AVT 310
D 204 SSSGSKIGDSGPPIPR-----EPMLRWSNGRPFELKDEKLVRNS-----EPQSGSIQ 250
QY 311 LD-KGKLQAPDNPALNTLLKQTLGKDTQHYLAHSSDGSQHLLLDNKGHLFDIKSTA 369
D 251 LDAGKPDFTNTPTGLAPLLDSILATPKQTYLAHQSKDGVGHQQLQANGHLLHLAQDD 310
```

```
QY 370 TSYSYLNHSHFC-EIKGKLAQAGTGSVSDVGKSGKISLGSQTSQSHNKTMLSQFBAHRS 428
D 311 SSLAVIRSSNEALLIEGKPPA----VKMEREDGNIHI--DTASGRKIQ-ELPGKAHIAH 363
QY 429 LTGTWQHPAGAPQGESIRLHDDKIHLHPGLGWQASD--KDT-HSOLSHQADQKLYA 485
D 364 ITNV-----LLSHDGERMYRVEDRLYQFDPISTRWKIPEGLEDTAPNSLSTGNGSVYA 417
QY 486 LKDNRTTLONLSDNKSSEKLVOKIRSYVDQGOVAILTDTQGRHKWSIMPSLDASPESHI 545
D 418 KSDDAV--DLSPMPMPHVEEDLQGSFVAPNRAALLS---GKTQAILLT--DMGP---- 468
QY 546 SLSLHFADAHQGLLHGKSELAQSVAIHSHGLVWADSEGLFSAAIPKQGGDNELMKAM 605
D 469 VIGGLTPKTKGLELDGGAQAAGVLSGDKLFIAQTQRLYSAD-RSAFEGDDPKLKM 527
QY 606 POHALDE----HFQHDHQISGFFHDDHQLNALVKNFRQOHCAPLG-NDHQFHPGWNLT 660
D 528 PEQANFOLEGVPLGCHNRVTGFIINGDDGCVHALIKNRQETHSHALDEQSSKLOSGWNL 587
QY 661 DALVIDNQLGHLHTNPPEP---HEILDHGLSLAQEGKLHVFQDLTKGWTGAE--SDCKQ 716
D 588 NALVLNNRGL--TWPPPTAADRLNLDRAGLVGLSEGRIORWDATPECWKDAGIKDIDR 645
QY 717 LKGLDGAAYLLKXGDEVKRLINQSTSIKHGTENVFLPHVRNKPPEGDALQGLNKDK 776
D 646 LQRGADSNAYVLKGGKLHALKIAAEHPNMAFDRTALTAQTAARSTKVMGKEIEGL--DDR 703
QY 777 A-QAMAVIGVNYKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISELGDHIVDHQON 835
D 704 VIKAFAMVSNKRFVALDD-----QNKLTASHKDKHPVTLDIPGLEGDIKLSLDEKHN 756
QY 836 LYALTTHEGEVHPQPREAWQN--GAE--SSSWHKLALPOSESKLSLDSHSHKPIATPED 891
D 757 LHALTSTGGLYCLPKEAWQSTKLGDQLPARPTVALPGQ--PVKALFTDNDNVLSAQIED 815
QY 892 GSQH---QLKAGGWHAYAAPERGLAVTSGTSQTVFNRLMQGVKGKVIPIGSGLTVKLSAQ 948
D 816 AEGKGLMQLKAGQQRFP---EQRP--VEENGLNDVHSRITGSKNTWRIPKTLGLTLRMDVN 870
QY 949 TGMGTGAEGRKVSXKFSRIPAYAFNPTMSTPRIPKNAAYATQGWQQRREGKPLPYEMQG 1008
D 871 TFGRSQVGEKSKAST-SEFIRANIYKNTAETPRWKVNVGDHIOHRYQGRGLGKEVYETES 929
QY 1009 ALIKOLDAHNVNRHNAPOP-----DLQSKLETLD--LGHGGAELNDMKRPREDELEQSATR 1061
D 930 MLFKQLEH--IHESGGPRPARGQDLKARITALEAKLGPQGAATLVKELETLELENHSYT 987
QY 1062 SVTVILG-----QHOGVLKSNGEINSFKEPSP-----GKALVQ---SFNVNRSGQDL 1104
D 988 ALMSIGQSYGKAKNLKQDDGILNQHGEL--AKFSVRMQFGKKLADLTGTLNFKSSGHDL 1044
QY 1105 SKSLQQAQVHATPPSAESKLSMLGHFVAGVDMHSHQGEIPLGRQRPDNDKTAITKSLI 1164
D 1045 VKELQDALTOVAPSAENPTKLLGLTKHQGLKLSHQKADIPLGQRDRASEDHGLSKARLA 1104
QY 1165 LDTVTIGELHELADAKLVSDHKPDADQIKOLRQOQFDTLREKRVESNPVHYTDMGFTHN 1224
D 1105 LDVLTKSLGALLDQV-----EQLPQSDIPLQKKLATLDTVTYGENPVKVVDMGFTHN 1160
QY 1225 KALEANYDAVKAFINAFKXHHGVNLTTRTVLESQGSAAELAKLNTILSLDSG-ESMSF 1283
D 1161 KALESGYESVKTFLKSPKADHVASVNMRAATGSKDOAELAGKPKMLKQLEHGDDEVL 1220
QY 1284 SRSYGGGVSTVFPVTLKSKVPVPIPGAGITLDRAYNLSFSRTSGGLN---VSFGRGGV 1340
D 1221 QRSYGVNLTPTPFIILADK--ATGLWPTAGATGNRYILNAERCEGVTVLVISEGA-GNV 1277
QY 1341 SGNIMVATGHVWPYM-----TGKKTSGAGNASDWLSAKHKLISPLRIGAAVSGTLQGTLO 1395
D 1278 SGGF--GAGKDYWPGFFDANNPAKSVVDGN-----NRTLTPNFRGLVDVTATVAASQR 1328
QY 1396 NSLKFKLTELDELPGFIHGLTHGTLPALQLKGIEHQMKQSKLTFESVDTSANLDRAGI 1455
```


Db 1329 AGVVNVPEBIDDAFVDDLPQGLNQLVKKAVDHSVEARRFNFDTAGGTADIRAGI 1388
 QY 1456 NLNED---GSKPN--GVTARVSAGLSA--SANLAAGSRERSTTSGQFGSTTSASNNRPTF 1508
 Db 1389 NLTEDRPDNPADPNSDSFSVAVRGGPAANIITVNIIMTYDYSLTQKNDKTELKGGKRNPRF 1448
 QY 1509 LNVGAGANITAAALGVAHSTHGGKPV--GIFPA--FTSTNVSAAL--ALDNRITSOSIS 1561
 Db 1449 LNNVTAGGQRAIGGSHAP--TGTASAGPTPASCATAANNLGGALNPFVSNRTVKRIK 1507
 QY 1562 LELKRAEPTVNSDISELTSTLGHKFDSDATTKMLAALKELDDA-----KPAE-----QLHI 1612
 Db 1508 FRYNVAKPIITTEGLSKLKGLEAFDNTTKAKLAELADPLNARYTGGKPDVITQAOLDG 1567
 QY 1613 LQOHFS---AKDVGDVEYAEVARNLKLVIROQA--ADSHSMELGSAHSTTYNNLSRIN 1667
 Db 1568 LEELFADIPPKD--NDKQYKALRDLKRAAVEHRASANKHSV--MDNARFETSNTNLSGLS 1624
 QY 1668 NDGIV-ELLKHFDPAALPASSAKRLGEMMNNDPALKDIIKQLQSTPPSSASVSMELKDGL 1726
 Db 1625 SESLITKINGSVRDASAP--GNATRVAFEMRQDPKLRAMLKEMEGSIGTLARVLEPKUSL 1683
 QY 1727 REOTEKAILDGKVRBEVGVLFQDRNNLRVKSVSQSVKSEGFNTPTALLIGTSNAAM 1786
 Db 1684 VDKIDEGSLNGTWTQSDLSMLEDRENMRIKRLVVFHTATQAEFTSTPTPLVSYNSGANV 1743
 QY 1787 SMERNITGINFKYQDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838
 Db 1744 SVTKTLGRINRVYGAQDQDKPIGYTFDGLGRPSASLKEAAGDLKKEGLEMK 1795

RESULT 5

O66101 PRELIMINARY; PRT; 1795 AA.
 AC O66101;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Avirulence protein (AVRE).
 GN AVRE.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=98115919; PubMed=9448330;
 RA Bogdanov A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
 RA Conlin A.K., Collmer A., Beer S.V.;
 RT "Homology and functional similarity of an hrp-linked pathogenicity
 locus, dmpE, of Erwinia amylovora and the avirulence locus avrE of
 Pseudomonas syringae pathovar tomato.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=98422476; PubMed=9748456;
 RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
 RA Collmer A.;
 RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
 to harpins and pectate lyases and can elicit the plant hypersensitive
 response and bind to pectate.";
 RL J. Bacteriol. 180:5211-5217 (1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=20243785; PubMed=10781092;
 RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
 RA Pernicki-Ocwleja T., van Dijk K., Collmer A.;
 RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
 mosaic structure composed of a cluster of type III secretion genes

RT bounded by exchangeable effector and conserved effector loci that
 RT contribute to parasitic fitness and pathogenicity in plants.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861 (2000).
 DR EMBL; U97505; AAC06134.1; -.
 DR EMBL; AF232004; AAF71499.1; -.
 DR PIR; T30332; T30332.
 SQ SEQUENCE 1795 AA; 195351 MW; 25074534DE08D5F CRC64;

Query Match 19.3%; Score 1824; DB 2; Length 1795;
 Best Local Similarity 29.0%; Pred. No. 6.5e-85;
 Matches 555; Conservative 366; Mismatches 744; Indels 246; Gaps 72;

QY 55 PRIH-----QSTAADGISAAHQKKSFLRGCLGTKKPSRSPQGGTTHSKCA-- 105
 Db 4 PSIHRTGSIQTPTVTPDARAATDLQER-----AEQRRSSHSLSVSG 47
 QY 106 --TLRDL--LARDDGETQHEAAPPDAAR-----LTRSGGVKRNMD-- 143
 Db 48 KRALKSVGKLFQSKAPQQAATPTAKNVKTPPPASNVATPRNKAREGFSNSPONT 107
 QY 144 -----MAGRPVMVKGSGEDKVPQTQOKRHQNLNFGQMRQTMLSKMAHPASANAGDRLQHS 197
 Db 108 RAPKWLILRNHPNQASSG-----AQTHEIPEAAPKRLRVRFDPQ-----DRLERS 155
 QY 198 PPHIPGSHHEIKBEPVGSSTKATTAHADRVEIAQEDDDSEFQHLHQORLARENPPQP- 256
 Db 156 PSYLDSDNPMTDEAVANATRF-----RSPDSHLOGSDGTRISMLATDPQPS 204
 QY 257 ---PKLGAT--PISARFQPKLTAVESVLBGTDTQSPKLPQSMKSGAGVTP--AVTL 311
 Db 205 SSGSKIGSDSGPIPPR-----EPMLWSNGRGREFELKDEKLVRNS-----EPQGIQL 251
 QY 312 D-KGKLQALAPDNPALNTLLKQTLGKDTQHYLAHASSSDGSQHLNLDNKGHLFDIKSTAT 370
 Db 252 DAKGKPDFSTFTPLGLAPLDSILATPKQYLAHQSKDGVGHQLQANGHFLHLAQDSD 311
 QY 371 SYSVLHNSHPG-EIKGKLAQAGTGSVDGSKISLGSTQSHNKTMLSQPGFAHRSLL 429
 Db 312 SLAVIRSSNEALLTEGKKPPA---VKMEREDGNIHI--DTASGRKTQ--ELPGKAHTAHI 364
 QY 430 TGIWQHPAGAARPOGESIRLHDDKIHLHPGLVWQSD--KDT-HSOLSRQADKLYAL 486
 Db 365 TNV-----LLSHDGERMRVHEDRLVQDFDPISTRWKIPEGEDTAFNSLTGGNGSVYAK 418
 QY 487 KDNRTLQNLSDNKSSEKLVDKISYSYDQRGQVAILTDTFPRHKMSIMPSLDASPESHIS 546
 Db 419 SDDAVV--DLSSPFMPHVEVEDLQSFVAPDNRAALLS---GKTQAILLT--DMSP---V 469
 QY 547 LSLHFDADAHQGLHKGSELEAQSVAIHSHGLVVADESKLFSAPKPGDGNELKMKAMP 606
 Db 470 IGGITPKTKTGLELDGGKAAAVGLSGDKLFIADTQGRLYSAD--RSAFEGDDPKKLIMP 528
 QY 607 QHALDE---HFGHDHIOISGFFHDDHGQLNALVKNFRQOHACPLG--NDHQFHPGNLITD 661
 Db 529 EQANFQLEGVPLGHNRTVTFINGDDGVHALLKNRQGETHSHALDQSSKLSQSGNLTN 588
 QY 662 ALVIDNQLGLHHTNPEP---HEILDGMHGLSLALQEGKLHYFDQLTKGWTGAE--SDCKQL 717
 Db 589 ALVLNNRGL--TWPPPTAADRNLNDRAGLVGLSEGRIOQRWDATPECWKDAGIKDIDL 646
 QY 718 KGLDGAAYLLKQGEVVKRLINQSTSIKHGTENVFSLPHVRNKPEGDALQGLNKDDKA 777
 Db 647 QRGADSNAYVLKGGKHALKIAAEHPNMFADRNLTALQATARTKSTKVMKEIEGL--DORV 704
 QY 778 -QAMAVIGVNVKYLALTEKGDIRSFQIKPGTQQLEPQAQTLRSREGISGELKDIDHVDHKQNL 836
 Db 705 IKAFAMVSNKRFVALDD-----QNKLTAKSKDHPVTLDDIPGLEGDIKSLSDKXNL 757
 QY 837 YALTHEGEVHPQPEAWON---GAE-SSSWHKLALPOSESKLKLSDMSHEHKPTATFEDG 892
 Db 758 HALTSTGGLYCLPKPEAWQSTKLGDLQRLARWTFVALPGQ--PVKALFTNDNDVLQAQIBDA 816
 QY 893 SQH---QLKAGGWHAAYAAPERGLAVGTSGSQTFVFNLMQGVKGKVI PGSGLTVLKLSAQ 949

Db 817 EKGMLQKAGOWQF---EORP---VEENGLNDVHSRITGSGNKTWIRPKTGLTLMEDVNT 871
Qy 950 GGMTAEGKVKSSKESRIRAFNPTMTPRPIKAAVATQHGQWQEGKPLKPLYEMOGA 1009
Db 872 FGRSGVEKSKAST--SEFIRANIYKNTAETPRMKNVGDHQRHYQGRGLGKLVETESM 930
Qy 1010 LIKQIDAHNVHRNAPQ-----DIQSKLETLD--LGEHGAELIINDMKRPRDELEQSATRS 1062
Db 931 LFKQLEL---IHESGGRPPARGQDLKARITALEAKLPGQCATLVKELETIRLELHNSVTA 988
Qy 1063 VTVLGL-----QHOGVLSKNSGEINSEKPPSP---GKALVQ---SFVNRSGQDLS 1105
Db 989 LMSICQSGYKAKNLKQDGLNLQHGL---AKPSVRMQFGKKLADLGLTKMFKSSGHDVL 1045
Qy 1106 KSLQAVHATPPSABSKLSQKLGHPVSAGVDMGHQGEIPLGRQRPDNDKATLTKSRLLI 1165
Db 1046 KELQDALTOVAPSAENPTKLLGLTKHQGLKSHQKADIPLGQRDASEDHGLSVARIAL 1105
Qy 1166 DTVTIGELHELADKAKLVSDHKPDADQIKQLRQOQDITLREKRYESNPVKHYTMDGFTNKK 1225
Db 1106 DLVTLKSLGALLDQV---EQLPQSDIEPLQKLLATLRDVTYGENPVKVVTDMGFTDNK 1161
Qy 1226 ALEANYDAVKAFINAFKKEHGVNLTTRTVLESQGSABELAKKLNTLLSLDSG--BSMSPS 1284
Db 1162 ALESYGSVKTFLSKFADHAVSVNMRRAATGSKDQAEAGKFKSMQLKLEHGDDEVLQ 1221
Qy 1285 RSYGGGVSTVFTPLSKKVPVPVPCAGITLDRAYNLFSRSTSGLN---VSFGRDGGVS 1341
Db 1222 RSYGNLTPFILLADK--ATGLWFTAGATGRNVLNARECEGVTVLYLISEGA--GNVS 1278
Qy 1342 QNIMVATGHDVWYV-----TGKTSAGNASDWSAKHKISPLDIRIGAASVTGLQTLQN 1396
Db 1279 GGF--GAGKDYWPFGFDANPARSDVGN-----NRTLTPNFRGLVDVTATVAASQRA 1329
Qy 1397 SLKPKLTDELPGRIHGLTHGLTPAEILLQKGIHQMKQSGKLFPSVUTSANDLRAGIN 1456
Db 1330 GVFNVPDEDIDAFVDDDFEQLNPLQVLKKAVDHESYEAREFNFPLTAGGTADIRAGIN 1389
Qy 1457 LNEED---GSKPN--GVTVARVSAGLSA--SANAAGSRERSSTTSGQFGSTTSASNNRPTFL 1509
Db 1390 LTERDPNADPNSDSFSVAVRGFGFAANTVNLMTYDLSLTKQKNDKTELKGGKRRPRFL 1449
Qy 1510 NGVAGANLTAALGVAHSTHEGKPV---GIFPA--FTSTNVSAAL--ALDNRTSQSISL 1562
Db 1450 NNVTAGGLRAQIGGSTAP--TGTPASAPGTPASQTAANNLGGALNFSVENRTVRIKF 1508
Qy 1563 ELKRAEPTVNDISLSTLSTLKHFKDSATYKMLAKELDDA-----KPAE-----QLHIL 1613
Db 1509 RYNTAKPITTEGLSKSLGSGLEAFDNTTKAKLAELADPLNARYTGKPKDEVIQAQLDGL 1568
Qy 1614 QOHFS---AKDVVDGDERYEAVRNLKLVIRQQA--ADSHSMELGSAHSTTYNNLRINN 1668
Db 1569 BELFADIPPPKD--NDKOYKALRDLKRAVBRHSAKXHSV--MDNARETSTKNLSGLSS 1625
Qy 1669 DGIV--ELLHKHFDALPASSAKRIGEMMNDPALKDIIKQLQSTPFSASVSMELKGLR 1727
Db 1626 ESILTKIMSSVRDASAP--GNATRVAEFMRQDPKLRAMEKMEGSGITLARVLRPKDSIV 1684
Qy 1728 EQTEKAILDGKVGREEVGLFQDRNRLVKVSVSQSVSKSEGENTPALLIGTNSAAMS 1787
Db 1685 DKIDEGSLNGTQSDLSMLEDREMEIKRLVVFHTATQAEFTSTPTPLVSYSNGANVS 1744
Qy 1788 MERNIGTINFKYGODNTPRFTLEGGIAQANPQVASALTDLKKEGLMKS 1838
Db 1745 VTKILGRINFYGADQDKPIGYITDGLSRFSASLKEAAGDLKKEGFLKS 1795

RESULT 6

Q8RNY8 PRELIMINARY, PRT: 1160 AA.

AC Q8RNY8:

DT 01-JUN-2002 (TremBLrel. 21, Created)

DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Type III effector AvrEpa (Fragment).
OS Pseudomonas syringae pv. maculicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=59511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ES4326;
EX MEDLINE=21862332; PubMed=11872842;
RA Guttman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA Greenberg J.T.;
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant
Pathogen Pseudomonas syringae";
RL Science 295:1722-1726(2002).
DR EMBL: AF458405; AAL84254.1; --
FT NON_TER 1
FT NON_TER 1160 1160
SQ SEQUENCE 1160 AA; 126695 MW; C0B67EE28860B991 CRC64;
Query Match 12.7%; Score 1198.5; DB 2; Length 1160;
Best Local Similarity 28.8%; Pred. No. 5.6e-53;
Matches 336; Conservative 225; Mismatches 476; Indels 128; Gaps 37;
Qy 313 KGKLQAPD-----NPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLDNKGHLFD 364
Db 58 QGTQLGADGKDPDSTFNTFCMAPLLSDILAKPQTVYLACQSQPDVQGHLLQSNRLLH 117
Qy 365 IKSTATSYVLHNSHPGEIKKLAQAGTGSVDSKSKISLGSGTQSH-----NKTM 417
Db 118 LAQDDSSLAIVRS-----TATLAVEANTPPALDMQVDDHIDTDPGRKS 164
Qy 418 LSQGEARHSLTGIWQHPCAGARP--QGESRLHDDKHILHLPGLVWQS---ADKQTHS 473
Db 165 LELPKLAHLAHITGVHKNKTGDRLPVREERLRIHEDRLYQDPTGFRWKAPEHMEIEAFN 224
Qy 474 QLSRQADGKLYALKNDRTLQNLSDNKSSEKLVDKIKSVSDQRGQVAILTDPGRHKMSI 533
Db 225 SLATGNGSIVAQSD--VVVTLSPFPMPHVEKELTVESVAPDNTAALLS---GKETAQV 280
Qy 534 M-----PSLDA--SPESHISLSLHFADAHQGLHKGSELEAQSVAISHGRLVVADEGKLF 587
Db 281 LLDTMSPVIGSLTPKTKALELD-----HGLA--QAVAIGLSAKKLFVADNOGRLY 329
Qy 588 SAAIPKQDGNELKMKAMPOHA---LDEHFGHDHQLSGFFHDDHQLNALVKNFRQOH 643
Db 330 SAD--RSAFERNEPTRLMPERTHYTLAQAMGSHNSVTGFINGGDGRVHALIKNRQGEIH 388
Qy 644 ACPLGND--HOPHPGNLTDALVIDNQLGLH--HTNPPEHILDMGHGLSLAQEGKLYHFD 701
Db 389 SHALDEQASKLESQNLNLTALVLDNTLGLNMPATAPEPANRLNDRAGLVGSGGRIQWMD 448
Qy 702 QLTGKWTGAE--SDCKQKKGLDGAAYLLKQGEVRLNI--NQSTSSIKHGTENVSFLPHV 758
Db 449 ATPQCKWDAGIKDIDRLQRGADSNAYVLKGGKLLRLTVTPHPNQAFDHT--ALAQTAR 506
Qy 759 RNKEPFGDALQGLNKDKQAQAMAVIGNVKYALATEKGDIRSFQIKPGTQQLEREAQTLRS 818
Db 507 STRVEMKEIEGL--EDRVITAFAMVSDTRFVALDDKNLLT-----VHKQKHPPVPMDF 558
Qy 819 EGISGLKDIHVDHKNLYALTHTHEGVPHOPREAWQ---NGAESSSWHKLALPQSSKL 874
Db 559 PGLGDIKELSLDEKHNLYARTSTGGFLCFLPKTAQANKLALLEGKWTVPATQGO--PV 617
Qy 875 KSLDMSHEHKPIATFEDGSQH---QLKAGGMHAYAAPERGPLAVGTSGSQTFVNLQGV 931
Db 618 KALFSDNDNLSAQIEDAPEQGLMQLKQGTWQTF---KQRP--VEENGLNDVHSKRKN 672
Qy 932 XGKVIKPSGLTVKLSAQGTGKMGTAEGKVKSKSERIRAYAFNPTMTPRPKAAATQ 991
Db 673 KTWRIPTGTLVTRTDVNLGAGGVKERNRPSA--GELWRANVYKPDGKVPFRMKNIGNHIQ 731


```
Db 1126 ANFAPVNVIALGSLGSGBP--SVAGFALFELSVVMVKFDRGESQAFSFGFKHPEPVTQSI 1183
Qy 1576 SELTSLGKHGKDSATTOMLAALKELDDAKPAEQHILQOHFSAKDVVG--DERYEAU-R 1632
Db 1184 DDLKESVSAY---SSTFKHDLAGIDQSCTTVDDQLATVQRFIAAHPMATKPDAYHAISQ 1240
Qy 1633 NLKULVIRQQAADSHMELGSAHSTTYNNLSRINNDGIVELLKHFDAAALPASSAKRLG 1692
Db 1241 SLDKLMTQODLVKNGLRQLASVSSVTRVGL---RDDG---RHAWLDDVAPANKA-AIE 1292
Qy 1693 EMMNNDPALKDIKLOLSTPSSASVSMELKDGLREOTEKAILDKGVGEEVGVLFQDBN 1752
Db 1293 QWLKDDPQLAQVNLQOHGEGSVKLMELKPVLRITERRHLAGDNTLEPLRQALSDDR 1352
Qy 1753 NLRKVSVSQSVSKSEGFNTPA-LLLGTSNAAAMSMERNIGTINFKYQGDQONTERRFTL 1811
Db 1353 NLRKVSMSLSYATQSHGMSIPAGANLSPSSAGLSYTKRVNADLEYGVDKDLRMNL 1412

RESULT 8
Q888Y1 ID Q888Y1 PRELIMINARY; PRT; 1957 AA.
AC Q888Y1;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Candidate type III effector HolPtor.
GN HOLPTOR OR PSPT00883.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Uettermack T., van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae";
RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE016858; AA054417.1; -.
DR TIGR; PSPT00883; -.
KW Complete proteome.
SQ SEQUENCE 1957 AA; 210393 MW; 2B94903E5E07BAE3 CRC64;

Query Match 3.4%; Score 325; DB 16; Length 1957;
Best Local Similarity 19.8%; Pred. No. 1.4e-07;
Matches 388; Conservative 249; Mismatches 710; Indels 616; Gaps 90;

Qy 208 IKPEPVGTSKATTAHADRVIAQEDDDSEFQQLHQRLARERENPPQPKLGIVATPISA 267
Db 2 VKVTSSTGFANPLSHADSVSPA-----NSPPQLPE-----PV-- 34

Qy 268 RFQKLTAVAESVLEGDTTQSPKQPMKSGAGVTPL-AVTLDKGK---LQIAPDNP 323
Db 35 ----HLVDLSRSSKGGMRN---RPHASLSNQ---VLELQAVPSQKGVVRSHADGE 83

Qy 324 PALNTLLKQTLGKTQHYLAHSHSDGSHLLDNKNGHLFDIKSTATSYSLV-----H 376
Db 84 SVINA-----WLAKRPSVQSETS--LNDNGKL--VRYTPVNHEPLAPRNEAFF 127

Qy 377 NSHPG-----EIKGLAQAGTGSVSDGKSLSGTSQSHNKMTLSQPGEAHRSLLT 430
Db 128 TSVFGMLMAVITVPEMEHGISGDITADAVARLIA-----EPPIGLLT 170

Qy 431 GIWHPAGAAAPQGESIRLHDKIHLHPGLVWQSAADKTHSLSQASQADGKLYALKDNR 490
Db 171 GIWQSSHDRAVLERGGV-VHTANMEERWAPLTPGINPREPLRMAGLQADGGVY-LHNGS 228

Qy 491 TLQNLSDNKSSEKLVVDKTKSVSVQDQGOVALITDTPGRHKMSIMPSLSDASPEHSISLSH 550
```

```
Db 229 QLWRL-----TETAAESVTTENLPEGAAVRI-----GAGGEVH----- 261
Qy 551 FADAHQGLH-----GKSELEAQSVAI-----SHGRLVVADSBGKLPFS 588
Db 262 --GLHEGALHNGISRPTELWRPKAGAFGRQSPARPVDLLPLPGGTAAIILDDKGRIVH 319
Qy 589 AAIKPKQG--DGNELKMKAMPQHAIHDEHFGHDQIS--GFFHDDHGLQNALVKNFNQQAHA 644
Db 320 ADLKTGTGAVEAHLKLPA-----DFAQKGWAVTAMGLSRDDTVHL--MLQDQNGRRNSL 372
Qy 645 CPLGNDHGFHCPGNUTDALVIDNQGLH-----HTNPEPHILDMGHLSLALQEG 695
Db 373 QRAPEALFRFAYLLDRELLLYTEGLHVPSEAAVQSHVQLDGH--AOLGHI-----DG 424
Qy 696 KLVFDQLTKGW-----TGAE-----SCCKQ-LKKGLDQ-- 723
Db 425 VLYHKAAPDQWERLKQSGGEPLTGLTALYSPLGFIORPKVPFALVGDARQVVELKLEGR 484
Qy 724 AAYLLKDGVEKR-----LNINQSTSIKHGTENVFSLPHVRNKPEPDALQGLNKDKKA 777
Db 485 TSWLPSDAELPRHPAGGFLAVIPDITVALRTSPIAQF-----DEPV 524
Qy 778 QAMAVIGVNYKYLALTEKDIRSFQIKPTQOQLERPAQTLRE-----GISBELKDIHVDH 832
Db 525 QALAVHGNRRVVALTDSGRMAADATPARLL-----PTLQRFIAIATAVGLNDQLLVHHPH 580
Qy 833 KQ-----NLVALTHEGEVFHQPREAWQNGABSSWH-KL 865
Db 581 SQRQLKLSAKDDWEPVPIILGLVHPSSLRATRTGQIQVLQGENWHITLPSMTSHDNQ 640
Qy 866 ALPOSESKLSLSDMSHEHKIATFEDGSOHLKAGWHAYAAAPERGLAVGTSGSQTVFN 925
Db 641 RLP---ARVKP-EPEGDSEANFLAGSNALANQQAASRISTPHEHDAVVTTLACTANN 696
Qy 926 RLMQGVKGVIPGSGLTIVKLSAQ-----TGMWTGABGRKV---SSKFSERIRAVAFNPTM 977
Db 697 -----PLTWASSLOAVDVTTRAQVGALARDVVGAANSTWRAHMTLGVVL 742
Qy 978 -STPRPIKNAAY--ATQHGQGR--EGLKPLVEMOGALI---KQLDAHVRHNAPQPD 1027
Db 743 PPTQEKRLASFHNEAKQYTSKGILFEHLPSLAQVRVASAVGSDSGRFGSLHQOQT-- 800
Qy 1028 LQSKLETLDLGEHGAELINDMKRFRDELEQSATRSVTVLGQHGVLI-----KSNGEI 1079
Db 801 --RLTLTREGKLEA-LLRDLRK-----IGFHEGVIMGDMGSDSASHGLV 841
Qy 1080 NSEFKPS-----PGKALVGSFNVNRRSGDLSKLSQQAQVHATPPSA----- 1119
Db 842 STTSTPTFLAELWRRQHSRVDKAL--SSAGLSRS-EDIFPDNLNLSINALAGGAALNDRM 899
Qy 1120 --ESKLQSMI---GHFVSAGVDMSHQKGEIPLGRQRPNDKTKALTKSRLILDVTTIGE 1172
Db 900 SERAEALLSVLCEVSEKMMRAGVRLPADDGVSVDASHQAP-----YGLRTAGL 947
Qy 1173 LHELADKAKLVSDHKPDADQIKQLRQOQFDTLREKRYESNPVKHYTDMGFTNKALEANYD 1232
Db 948 IAGLVVDYDALIS--STDAQALE-----MAERLQDARLAALCKLGLSGWGLAAFDD 997
Qy 1233 AVKAFINAFKEHGVNUTTRTVLESQGSABLAKLKNTLLSLDSEGSMSFSRSGVGS 1292
Db 998 VVITTF-----REQISLPGSARRTQLKNLGLPDDAAPEDEMAARM----- 1036
Qy 1293 TVFVPTLSKKVPVPIPCAGITLDRAYNLS--FSRTSGGLNV--SFGK-----DGG 1339
Db 1037 -----SOLLLD-LFNRRSTFTQSRGELRGLSGSADMKHLNFAVSG 1077
Qy 1340 VSGNIMVATGHDVMPYMTGKTSAG-----NADSWLSAKHKISPOLRIGAAVSGTQGG 1392
Db 1078 VTGALQVLG--VERIGDKDGDAGLVAFFVRHAKASVSATSGIGIDFKPGPGTGGRV-- 1133
Qy 1393 TLQNSLKFKLTDELPQFIHGLTHGTLTPAELLQKQIE--HMQKGS----- 1437
Db 1134 -----IDSRPGRSNMNSTWGGST-----NLGISGAYOHGQGAIVLAPSTISDFV 1177
```

QY	1438	KLTFSV---DTSANLDRAGINLNEDGSK--PNGVTARVSAGLS-----	1476
Db	1178	RLFDVNPDDTQ--ILRTGVNGSGIGLDLPETNVASGVNSVSPFSLSKYQVKPQT	1235
QY	1477	ASANLAAGSRERSTTSQGF--GSTTSA-----	1508
Db	1236	ADAAVSGPNNRRSTAGSLSVGGTAQAGAHWGOMELHLDHAWADIIGLEFQGRITDFNJEF	1295
QY	1509	INGVAGAGANLTAALG-----VAHSTHEG--KPVGIFPAFTSTNVSAALADNR-----	1555
Db	1296	NEGMLGGALSALGDNPNFKLINASTGNNGNQLAGIRVASSDVQLPTDAVVDKRRGPFL	1355
QY	1556	--TSQSISLEKRAEPVTSNDSIELTSLTGLGHFKDSATTKMLAALKELD-DAXPAEQLHI	1612
Db	1356	STASVKRFTFDEVAEPVTAGESQMRQLAKAFDN-----IAELGALDYTPRGERIAT	1410
QY	1613	LOQHSFAKDVVDGERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSINNDGIV	1672
Db	1411	IKQVIDRIQGAKEARVEAVGAMDGKALHQRQFDA-AREMSNAGNS-VVRASSEIERASIV	1468
QY	1673	ELIHL-----KHFDALP-----	1695
Db	1469	EMLHQLRQEQSAVQNHARAPGARVEFNLFGRSLEFVTFVFAIGHLGLGSKLNDUABLH	1528
QY	1696	NNDPALKDIILKQLQSTP-FSSASVSMELKDGLREQTEKAILDGVKVGREEV-----	1744
Db	1529	RKVPGLDQVMSLSQPLKPNQVRYFE---MRQARFANDALLAREQOASARALGLOG	1584
QY	1745	-----GVLFQDRNN---LRKVSVSQSVSKSEGFNTPALLLGTSTNSAAMSMERNI	1792
Db	1585	PSGSELNWRGVLDKIKTTPDLRYLAAIAVHNTDENPVTSRIGLPLLNVSATGATSHQLUFE	1644
QY	1793	GTINKYQGDQNTPRFTLEGGIAQNPQVASALTDLKKEGLE	1835
Db	1645	AEIQFRIYGLYDGLQGVLELAG---NRALQSPRALQQSGIQ	1683
RESULT 9			
Q8QX66 PRELIMINARY; PRT; 1702 AA.			
ID	Q8QX66	PRELIMINARY; PRT; 1702 AA.	
AC	Q8QX66;	AC	
DT	01-MAR-2002 (TrEMBLrel. 20, Created)	DT	
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)	DT	
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)	DT	
DE	Hypochemical protein RSP1281.	DE	
GN	RSP1281 OR RS05326.	GN	
OS	Ralstonia solanacearum (Pseudomonas solanacearum).	OS	
OC	Plasmid megaplasmid.	OC	
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;	OC	
OC	Burkholderiaceae; Ralstonia.	OC	
OX	NCBI_TaxID=305;	OX	
RN	[1]	RN	
RP	SEQUENCE FROM N.A.	RP	
RC	STRAIN=GMI1000;	RC	
RX	MEDLINE=21681879; PubMed=11823852;	RX	
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,	RA	
RA	Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,	RA	
RA	Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,	RA	
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,	RA	
RA	Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,	RA	
RA	Weissenbach J., Boucher C.A.;	RA	
RT	"genome sequence of the plant pathogen Ralstonia solanacearum.";	RT	
RL	Nature 415:497-502(2002).	RL	
DR	EMBL; AL646084; CAD18432.1; -.	DR	
DR	GO; GO:0046821; C:extrachromosomal DNA; IEA.	DR	
KW	Plasmid; Hypothetical protein; Complete proteome.	KW	
SQ	SEQUENCE 1702 AA; 181169 MW; 296517012D122445 CRC64;	SQ	
Query Match 3.2%; Score 302.5; DB 16; Length 1702;			
Best Local Similarity 19.2%; Pred. No. 1.7e-06;			
Matches 372; Conservative 243; Mismatches 758; Indels 569; Gaps 85;			

Db 871 LRGAIDRWLG-----ALAQIGVAAVLGPDLTLDAAAGRRKQRRPAKADDP 918
Qy 1119 AE--SKLQSMGLGHFVSAGVMSHOKGEIPLGRQDPNDKTALTSLRLDITVTIGELHEL 1176
Db 919 DDLPLKMSWDEFV-----RAE-----AQKPSPEAGAKAPVPFDPARLGDHRI 966
Qy 1177 ADKAKLVSDH-----KPDADQIKOLRQOF----- 1200
Db 967 VELMHTLSEHGKVPAPDLTGQDTRDRFAILTGAAGRHLTLDLATRLVTLPTQVAVIG 1026
Qy 1201 -----DTLREKRYENPVKHYTDMGTFTNKALANYDAVKAPINAFKEHGVN---ITT 1252
Db 1027 SDSAVEVLRSQ--VETDKVLRLAREFGFSNNWDAEAFWETAQTFKKEVWMPGSPFNKFAES 1085
Qy 1253 RTVLBSQGSAAELAKLKNLTLSDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVPCAG 1312
Db 1086 HAVRDASSPAEMALKPAEMAKGLNRSSTLFGIESKAGASVTSAGVLPKRVGAFGPAV 1145
Qy 1313 ITLDRAYNLSFRTS-----GGLNVSGRDGGVSGNIMVATGHDVMPYMTGKTSAGNAS 1367
Db 1146 LGVDRTTWGVERTADKLAEGPLVAFFVRQSN-----KGLTWGGGGM 1188
Qy 1368 DWLSAKHKISDPLRIGAAVSTLQGTLONSLKFKLTDELPCTFHGL-----THGTLTPAE 1423
Db 1189 DFKPLRKVLGPRLOGGAQASASVMH--GKGAAMLVAPDNIDEFARRLPDPASH---DPCR 1243
Qy 1424 LLOKGIHOMKGGSKLTFSVDTSANLDRAG-----INLNEGSKPENGVT 1469
Db 1244 LLELG-----INQGAIGLDMIEQQNLNSANAGLGGYADKIPGFPAHRHGGDS---STSA 1296
Qy 1470 RVSAGLS---ASANLAAGSRE-----RSTTSGQFGSTTSASNRRPTFLNGVGAGANL 1518
Db 1297 FOOTGLORGEVGVANWVGVRDFELKLOHAWEPISGY-----EYQGGRGWSANA 1344
Qy 1519 TAALGVASHSTHEGKPVGIFP-----AFT-----STNVS---AALALDNRTSOSISLELRA 1567
Db 1345 FASL-----VQOG---GLLPHVSDAFTLVLRSLNITLLGASVELSGVESFKRTLDMKTA 1395
Qy 1568 EPVTSNDISELTSTL-----GKHFQDSATTKMLAALKELDDAKPAEQHLILQ 1614
Db 1396 ARVTPQWESQAALAREVFPQATGHGDPHLKAIATLQAA-----K 1439
Qy 1615 QHFSKADVGDYEREAENLKKLVIRQO--AADSHSMELGSAHSTTYNNLSRINNDGIVE 1673
Db 1440 STWAAR--TEHERASFVDRAEQLLLDQDQASGRAMLPGAKIERNIPNFRSLVDTRKNS 1497
Qy 1674 LLKHFDALPASSAKRLGEMMNNDPALKDIIKQLQSTPFPSSASVSMELKDLRE----- 1728
Db 1498 KAHRSGLALMEA--AERAREAV--PGLADAMR-----AMSERDGVNDVRFVP 1540
Qy 1729 -----QTEKAILDGKVGREEVGVLFQD--RNLRVKSVSQSVSKSEGN-TPALLL 1778
Db 1541 QMDESYINAVNRLMEGLKSWAEFNTWARTVPAPYRLTEICAKDSNRSFTLNPLPL 1600
Qy 1779 GTSNSAAMSERNITGTFNKYG 1800
Db 1601 AFNDSAEVRSLEAEVHLRYG 1622

RESULT 10

Q8NWQ6 ID Q8NWQ6 PRELIMINARY; PRT: 9904 AA.
AC Q8NWQ6;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Ehb protein.
OS Ehb OR MW1324.
GN Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_taxID=196620;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMU; AP004826; BAB95189.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR002988; GA.
DR InterPro; IPR005977; Gpos_YsIRK.
DR InterPro; IPR006530; YD.
DR Pfam; PF01468; GA; 48.
DR Pfam; PF04650; YsIRK signal; 1.
DR TIGRFAMs; TIGR01643; YD repeat_2x; 8.
DR TIGRFAMs; TIGR01168; YsIRK signal; 1.
KW Complete proteome.
SQ SEQUENCE 9904 AA; 1068476 MW; CE86908F75EA56B2 CRC64;

Query Match 3.0%; Score 283.5; DB 16; Length 9904;
Best Local Similarity 18.6%; Pred. No. 0.0003;
Matches 396; Conservative 306; Mismatches 817; Indels 605; Gaps 94;
Qy 8 TEHKAHVHTAAHNPVGHGVALQQ--GSSSSSPQNAASLAAGKNGKMGKPRIHQPSAAD 65
Db 2543 TQYNAIHNAQQQINTAKTEAQVINDRATPQQVSDALT-----KVRAAQTKID 2592
Qy 66 GISAAHQKKSFSRLGCLGTTKFSRSAPQGGPQGTTHSKGATRLDRLARDGETQHEAAP 125
Db 2593 QAKALLQNKEDNSQ---LVTSKNLQSSVNVQVST--AGMTQQSIDNNAKKREAREIT 2647
Qy 126 DAARLTSSGGV-----KRRNMDDMAGRPMVKGSGEDKVPQCKRHQLNNFQMQRT 177
Db 2648 AQRVINDGATPQOISEKHRVDNALTNLQAQKNTADTHTEQAVQQLNRTG---T 2703
Qy 178 MLSKMAHPASA---NAGDRLQHSPPHIPGSHHEIKEEFVSGSTSKATTA--HADRV--- 227
Db 2704 TTGK--KPASITAYNNSMHALQASLTSAKNSANALIQKPIRSVQEVQIATNVRNREL 2761
Qy 228 -----ETAEEDDDSEFQOLHOORLARENPQPKLGAVATPI SARFQKLTAV----- 276
Db 2762 TQAINQLVPLADNSAL-----RTAKTKLD---EINKSVTTDGMTQSSIQAVENAKRA 2811
Qy 277 -----AESVLEGTDTTQSPKQPQ-----SMLKSGAGVTPPLAVTLDKGLQLAPD 321
Db 2812 GQTESTNAQVNNINGDATDQOIAEBKVEKYNLSKQAIAGLTPLDPLAPLOFTAKTQLOND 2871
Qy 322 -----NPPALNTLLK--QTLGKDTQHYLAH-----ASSD 349
Db 2872 IDQPTSTTGMTSTSI AAFNEKLSAARTKI QEIDRVLASHPDVATIRQNVTAANAATAKD 2931
Qy 350 GSQHLLDNKGHLDPDKS-----TATSVLHNSHPGETKGLAOGTGSVSDVSKG 402
Db 2932 QARNGLTVDKAPLENAKQLQHSIDTDTSTGMTQDSINAYNAKLTAARNKIQQIN----- 2987
Qy 403 KISGSGTQSHNKTMLSQGEAHSLLTGIWHQFAGARPQGESIRLHDDKHILHPLG 462
Db 2988 QVLAGSPTEVQINTNTSAANOAKSDL-----DHARQALTP-----DKAPLOTAKTQ 3033
Qy 463 VWQSAKDT-----HSQLSRQADGKL-----YALKDNRTLQNLSD-----NKSS 501
Db 3034 LEQSINQPTDTTGMTTASLNAYNQKLAARQKLTETINQVLNGNPTVQVINDKVTETANQAK 3093
Qy 502 EKLVDKIKSVSDQGVQVAILTDTPGHKKMSIMPSLDASPEHSLSLSHFADAHQGLLHG 561
Db 3094 DQLNTARQGLTDRQPALTTL-----HGAS---NLQAQQNNFTQIINAQAHALETI 3144
Qy 562 KSELEAQSVASHGRVLVADSEGLFSAAIPTKQDGNELKMKAMPQALDEHFHGHQHS 621
Db 3145 KSNITALTAMTKLQDSVAD-----
Qy 622 GFFHDDHGQNLALVKNNFRQHQHACPLGNDHQFHPGWNLTDALVIDNQGLHHTNPPEHI 681

Db 3165 -----NNTIKSQNYVDATP-ANKQAYDNVNAKGVIGET-----TNP-----T 3203
Qy 682 LDMCHGLSLALQEGKLYHFDLTGKTGABSDCKQLKKGLDGAAYLLKQGVKRLINQS 741
Db 3204 MDVNTYNQKA-----ASYKSTKDALDGOONLQRAK-----TEA 3236
Qy 742 TSSIKHGTENVFSLPHVRNKPEPDALQGLNKDDKAQAMAVIGVKNYLALTEKGDIRSFO 801
Db 3237 TNAITHASD-----LNAQKNALTOQVN-----SAQNVHAVN-----D 3269
Qy 802 IKPGTOQLERPAQTLGREGISBELKDHYD-----HKONLY--ALTHEGEVPHOPREA 852
Db 3270 IKQTTQSL-NTAMTGLKRGVANHNVQVSDYVNVADTNKNDYNNAYNHANDII----- 3322
Qy 853 WQNGARSSWHKALAPQSESKLSLSDMSHEHKFIATFEDGSQHLKAGGWHAAAPERGP 912
Db 3323 --NG--NAQHPVITPSDVNNALSNVTSKEH-----ALNGEAKLNAKQAEAN 3364
Qy 913 LAVTSGSQTFVFNRLMGVKGKVIPIGSL-----TVK-----LSAQGTGGMTCABGRK----- 959
Db 3365 TALGQLNN--LNNARQNLQSQINSAHQIETVNTIKQATNLNSAMGNLRQAVADKQOVK 3422
Qy 960 -----VSSKFSERIRAYAFNPTMSTPRPKNAAYATQHGQGBEGL 1000
Db 3423 RTEDYADADTAKQAYNSAVSS--AETIIINQTTNPTMSVD--DVNRATSAVTSNKNALNGD 3479
Qy 1001 KPLYEMOGALIKOLDAHNVHNAPOPLQSKLETLD-----LGEHGAEL----- 1044
Db 3480 EKLAQSKTDAARAIDALPHLNNQAKADVKSINAASNIAGVNTVQOGTDLNTAMGNLQ 3539
Qy 1045 -LND-----MKRFRDELEQSARSRTVLQHQGLV--KSNGEINSEFPKSPGKALVQSFN 1096
Db 3540 AINDEQTLINSQYQDATPEKKTAYTNAVQAADILNKSNGQ--NKTKDQVTEAMNQ--LN 3596
Qy 1097 VNRSGDLSKLOQAVHATPPSAESKLSQMLGHFVSVAGVMSHOKGPIPLGRQDPNDKT 1156
Db 3597 SAKNLDGTRLLDQA-----KQAKQQLNNMT-HLTTA-----QKT 3631
Qy 1157 ALTKSRLILDTVIGELHELADKALVSDHKPDADQIKOLROQFDTLREKREYNPNVKHY 1216
Db 3632 NLTNQ--INSQTTVAGVHTVQSN-----NTLDQAMNTLRQSIANKDATKAS 3676
Qy 1217 TDMGFTHNKALEANYDAVKAFFINAFKEHHEGVNLTTRTVLESQGSALAKKNTLLSLD 1276
Db 3677 EDY-----VDANDKQAYNNAVAAETIINANSPEMNPSTITQKAEQVNSKTA 3729
Qy 1277 SGESMSPSRYSYGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSPSRITSGGLNVSPGR 1336
Db 3730 GDENLATAKQN-----AKTYLNTLT-----SIT-DAQKNLISQISSATFVS-GV 3772
Qy 1337 D-----GGVSGNIMVATGHDVMPYMTGKTSAGN-----ASDWLSAKHK 1375
Db 3773 DTWKQNAQHLDAQMASLQSGINNESQVSKSEKYPEDADTNKQOEYDVAITAKALNKHQ- 3831
Qy 1376 ISPLD-----RICAAYS--GTLQGTQNSLKFELDELPFGTHGLTHGTLTPAEILLQKIE 1430
Db 3832 -GPNTAQNAVEAALQRYNTAKDALNGDAKLIQAQNAKQHLGTLTH--ITTAQ--RNDLT 3886
Qy 1431 HQMKQSGKL--TFSVDTSAN-LD-----LRAGIN-----LNEGSKPQNGVTVARV 1471
Db 3887 NQISQATNLAVGESVQKNANSGLDAMGNLQTAINDKSGTTLASQNFLLDADEQKNAYNQAV 3946
Qy 1472 SAGLSASANLAAGSRERSTTSQGFSTTSASNNRPTLVG-----GAGANLTAALGVAH 1526
Db 3947 SAA--ETIILKQGTGNTAKTAVEQ-----ALNNVNNAKHALNGTQNLNNAKQAAITANGASD 4002
Qy 1527 SSTHEGKPVGIPPAFTSNVSAALALDN-RFSQISLELKEAEFVTSNDISELTSTL-GK 1584
Db 4003 LNOHQKDAL-----KAAQANGAQRVSNQADQVRNATELNTAMGTLKHAIADKNTLASS 4055
Qy 1585 HFKDSATTKMLAALKELDDAK-----PAEQHLILQOHFSAK-DVVGDER----- 1627

Db 4056 KYVNADSTKQNAVTTKVTNAEHIISGTPVTVPVTTPTSEVTAANQVNSAKQELNGDERLVA 4115
Qy 1628 -----YBAVRNL-----KKLVIRQQAADSHSMELGSAHSHTTYNNLSRINNDGIVELLHK 1677
Db 4116 KQNAVTAIDALTQNTPOKAKLKEQVQANTLD-----DAMNSLOGAINDKDATLRNQ 4168
Qy 1678 HFDALPASPASAKR-----LGEMMNNDPALKDIIKQLOSTPFPSSA----- 1716
Db 4169 NY---LDAESKRNAYTQAVTAAGILNKQTGNTSKADVNDALNAVTRAKAALNGAENL 4225
Qy 1717 -----SVSMELKDGLEOTEKA-----ILD-GKVGREEVGVL-----FQ 1749
Db 4226 RNAKTSATNTINGLPLNLQLOKNDLKHQVEQANVAGVNGVKQKGTNTAMGALRTSIQ 4285
Qy 1750 DRNLRKVSQSVQSVKSEGFNTP-----ALLIGTSNAAAMGWNIGTINFKYGDQDN 1804
Db 4286 NDNTTKTSQNYLDASDSKNKNYNTAVNNANGVINATNPP---NWDAN--AINGMANQVNT 4340
Qy 1805 TPRRFTLEGGIAQANPOVASALTD 1828
Db 4341 TKAALNGVQNLQAQAKTNATNTINN 4364

RESULT 11
Q9LAW1 ID Q9LAW1 PRELIMINARY; PRT; 57 AA.
AC Q9LAW1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE Hrp-secreted pathogenicity/avirulence protein DsPE (Fragment).
GN DsPE.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=Ba246;
RA Kim J. F., Laby R. J., Beer S. V.;
RT "Comparison of the hrpN-flanking regions of two Erwinia amylovora
strains with different host specificity.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RD EMBL; AF083620; AAF63403.1; -.
FT NON TER 57
SQ SEQUENCE 57 AA; 5788 MW; AFFOCA36311E4BE1 CRC64;

Query Match 3.0%; Score 283; DB 2; Length 57;
Best Local Similarity 98.2%; Pred. No. 5.8e-08;
Matches 56; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELKSLGTEHKAHVHTAAHNPVGHVVALQGGSSSSSPQNAASLAABGKNGKMPRI 57
Db 1 MELKSLGTEHKAHVHTAAHNPVGHVVALQGGSSSSSPQNAASLAABGKNGKMPRI 57

RESULT 12
Q99U54 ID Q99U54 PRELIMINARY; PRT; 6713 AA.
AC Q99U54;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Eha protein.
GN Eha OR SA1267.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K. I., Nagai Y., Lian J., Ito T., Kanamori M.,

QY 979 TPRPIKRAYATQHWQREGIKPLYEMQCALIKOI-----DAHNVTHNAP 1024
Db 854 ----IHNAGATIEAGKRLRGYVEKLHNTNEHLKTLQVETGREHIVDYFAFGHELLREGT 909
QY 1025 OPDL-----QSKLETLDLGEH-----CAEILNDWK 1049
Db 910 QHELGSVYNDESRLRTPDGAHWHYDYKVTQKTQVTTAPAKIISGNDITDQK 969
QY 1050 R-FRDELEQSASTRSVTLGQHG-----VLKNGEINSBFKPS-PGKALVQSFN 1096
Db 970 EVENTDSQIAGNLIVQTEKDLHNEQTFGEKKVSENGKLHSYREKHKR---DSNG 1026
QY 1097 VNRSCDLSKSLQAVHATPSAESKLSMLGHFVSAGVDMHQGEIPLGRQRPDKT 1156
Db 1027 HSEQNTLPETIRNISLGSFAYESHKALSHAPSQGTLPQSNQ---ISLPTYSNFT 1083
QY 1157 ALTKSRL-ILDTVTIGELHELADKAK-----LVSPHKPDA---DQIKQLRQQDFTLREKR 1207
Db 1084 PLPSSSLIINPVNGYLVETDPRFANYRQWLGSDYMLDSLKLDPNHLKRLGCGYYEQR 1143
QY 1208 YESNPVKHYTMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSSELAKK 1267
Db 1144 LINEQIAELT--GHRRLDGYQNDDEQFKALMDNGATAARSNNLSVGIAL-----SAEQVAQ 1197
QY 1268 LKNTLLSLDSGESMSFSRSYGGGVSTVPPLSLKKVPVPIPGAGITLDRAVNLFSRST 1327
Db 1198 LTSIDVILVQKE-----VKLPDGGTQTLVPOVYVRVKNQDIDKGAL-----LSGSNTQ 1247
QY 1328 GGLNVSGRGGVSGNIMVATGHDVPMYTGK---KTSK-----GNASDWLSAKHKIS 1377
Db 1248 INVSGSKNSGTIAGNALLINTDLNIGRIHQAQSAVATQDINNIGMLSAEQTLL 1307
QY 1378 PDLRIGAAV-SSTLQGTQLNSKFLKTEDELPG-FIHLGTHGTLT-----PAELLQKIE 1430
Db 1308 --LNAGNINSQSTTASSQNTQSGSTYLDRMAGIIVITCKEKGVLAAQAGKDIINIAGQIS 1365
QY 1431 HOMQGSKLTVSDTSANL-----RAGINLNEGSPNGVGTARVSAGLSASA----- 1479
Db 1366 NQSBQG-QTRLQAGRDINLDTVQTSKHOATHFDADNHVIRGSTNEVGSSIQTKGDVITLLS 1424
QY 1480 --NLAAGSRERSTTSGFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSTHEGKPVGI 1537
Db 1425 GNNLNKAAAEVSSANGTL--AVSAKND-----IN-ISAGINTT---HVDDASKHTGRSGG- 1473
QY 1538 FPAFTSTNVSAALADNRTSISLELKRBPVTSNDISELTSTL-----GKHFK 1587
Db 1474 ---GNKLVITDKAQSHHETAQSTFEGKVVLQAGNDANILGSNVISDNGTQIQAGNHVR 1530
QY 1588 DSATTKMLAALKELDDAKPAQLHLQHFSAKDVGVDERYEAVERNKKUVIROQAADSH 1647
Db 1531 IGGT-----QTSQSETYHQTKSGLSAGIG-----FTI---GSKTN 1565
QY 1648 SMLGSASHSTYNNLSRINNNGTIVELLKHFDALPASSAKRLGEMM-----NNDPALK 1702
Db 1566 TQENQSQSNEHTGSTVGLKGDYI-VAKKYE-----QIGSTVSSPEGNTIYAAQ 1615
QY 1703 DIIKQLOSTPSSASVSMELKDXLREQTEKAILDKGVGREVG-----LFQDRNNLRV 1756
Db 1616 SIDIOAAHNKLSNNTQTYEQKGLTVAFSSPTD--LAQAIAVAQSSQVQSKND-RV 1672
QY 1757 KSVSVSQS--VSKSEGFNTPALIGTSNAMSERNITGTFNFKYQDQNTPRFT-LEG 1813
Db 1673 NMAAANAGWQAYQTKGSAQNLANGTNTAKQVS-----ISITYGEQON--RQTTQVQA 1723
QY 1814 GIAQNPQVASALTDLKEGLEMS 1838
Db 1724 NQAASQIQAGGKTTLIATGAEOS 1748

RESULT 14

Q931R6

ID Q931R6

AC Q931R6;

PRELIMINARY; PRT: 6713 AA.

DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Hypothetical protein ebhA.
GS EBHA OR SAV1434.
CN Staphylococcus aureus (strain Mu50 / ATCC 700699).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=213111952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus."
RL Lancet 357:1225-1240 (2001).
DR EMBL; AP003362; BAB57596.1; --
DR InterPro; IPR002988; GA.
DR InterPro; IPR006530; YD.
DR Pfam; PF01468; GA; 46.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 7.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 6713 AA; 722306 MW; BBCC0536AC341BF5 CRC64;

Query Match 2.9%; Score 273.5; DB 16; Length 6713;
Best Local Similarity 17.8%; Pred. No. 0.00051;
Matches 407; Conservative 322; Mismatches 904; Indels 651; Gaps 95;

QY 11 KAAVHTAAHNPVGHGVAL-----QGGSSSSPQNAASLAAGKNGKMPRIHOPSTA 63
Db 556 KAAVENALSVQTNKAGLNGNHNLEQAKSNANTINGLOHLLTAKOKLKQ-QVQQAQNV 614
QY 64 ADGISAHQKKSFLRGLGTFKFSRPAQOGPTTHSKGATLRDLARDDDGETOHEAA 123
Db 615 A-GVDTV--KSSANTLNGAMGTLSNI-----QNTATKNGQNYLD--ATERNKINYNA 664
QY 124 APDAARLTRSGVKKRRNMMDMAGRPVKGSGEDKVPQT--QKRHL---NNFGQMRTM 178
Db 665 VDSANGVINA--TSNPNMD-----ANAINQATQVTSKNDLGTNLTQAKQTA 712
QY 179 LSKMAHPASANAGDRLQHSPPHIGSHHEIKEEPVGSSTKATTAHADRV-----ETAQ 231
Db 713 TNAIDGATNLNKAOK-----DALKAQVTSQORVANVTSIQQTAN 751
QY 232 E-----DDSEFQQLHQRLARENPFPKLGVAATPISARFQPKLT--AVA 277
Db 752 ELNTAMGQLQHGIDENATKQTKYRDAEQSKKTAQDAVAAAKAILNKQTGNSDKAAV 811
QY 278 ESVLEGTTTQSPKLPKSMKSGAGVTPIAVTLDDKGLQAPNPPALNTLLKQTLGKD 337
Db 812 DRALQQTSTKDALNGDAKLAEEAKAAARQNLGTLN---HITNAQRTALEQIQIQAATVD 867
QY 338 TQVLAHHASS-DGSOHLL-----LDNKGHLPDIKSTATSYSVLHNSHPEIGK 386
Db 868 GVNTVTKNTNLDDGAMNSLQCAINDKATLRNQNYLDADESKRNWAYTQAVTAAG-ILNK 926
QY 387 LAQAGTGSVSDGSGKISLGSQTSHNKTMLSQPGEAHRSLS--LTGIVQHPAGAAARPOG 444
Db 927 QTGGNTSKADVNDALNAVTRAKALNCAENLRNAKTSATNTINGLPNTLQLOKDNLKHQV 986
QY 445 E-----SIRLHDDKIHLHPELVWQS-----ADKDTHSQLSRQA 479
Db 987 EQAQNVGVGVNGVKDQKGTNTLNTAMGALRTSIQNDNTTSTQNYLDASDSNNKNYNAVNA 1046
QY 480 DGKLYALKD-----NRT-----LQNLSDNKSSEKLVDKI-KSVSVQORQ 517
Db 1047 NGVINATNPNMDANAINDMANOVNTTKAALNGAQNLAQAKTN--ATNTINNAQDLNQKQ 1104

DR SMART: SMO0382; AAA; 1.
 DR PROSITE; PS00387; PPASE; 1.
 KW ATP-binding.
 SQ SEQUENCE 1877 AA; 202566 MW; A09BF57D255609BE CRC64;

Query Match 2.8%; Score 261.5; DB 4; Length 1877;
 Best Local Similarity 20.1%; Pred. No. 0.00025;
 Matches 355; Conservative 231; Mismatches 713; Indels 465; Gaps 79;

QY	31	GSSSSSPQNAASIAEAGKRGKMPRIHQPTSTAADGISAAHQKKSFSLRGCLGTGKKFSR	90
DB	20	GDECADEPRGAGRAAAAADGRMLPKRAKAPGGGGMAKASAAELKYFKSGSVDSR	75
QY	91	SAPQOGTTHSKGATRLDILLARDGTOHEHAAPDAARLTRSGGVKRRNMDDWAGRP	148
DB	76	VPGGPPASNLRLKOKSLTNLFUTDSEKLLQLYEPWS	121
QY	149	MVKGSGEDKVPTQCKRHLNNFQMQTMLSMAH	190
DB	122	LGVGSGKREAPLSK	168
QY	191	GDRLOHSPHIP-GSHHEIKEEPPVGSTKATTAHADRVETIAQEDDDSEFQOLHQRLARE	249
DB	169	G-----KPSRIPRGPYAEVK-----PLSKAP-----EAAVSEDGKSDDELSSKAKAQ	211
QY	250	RENPPQPKLGVATPISARQPKLTAVAESVLGTTDTTQSPFLPQSMKSGGAGVTPLAV	309
DB	212	KSGGFVPSAQGEERAFKVDPELVTVLGDLEQLLFSQM-LDPESQRRKTQNV	265
QY	310	TLDRGKQLQAPNPPALNTLLKOTLGKDTQHYLAHAHASSDQSO--HLLLDNKGHLFDIKS	367
DB	266	LD-----LRQNLLETM-----SSLRGSQVTHSLEMTCYDSDAN	300
QY	368	TATSYSVLHNSHP-----GEIKGLAQAGTGSVSDGSKISLGSGTGQSHNKTMLSQPG	422
DB	301	PRSVSSLSNRSPLSMWRYGQSSPRL-QAGDAP-----SVGSGCRSEGTPAWYMHG	349
QY	423	E-AHRSLLTGWOHPAGAAAPQGESIRLHDDKTHILHPELGVWQSAKD-PTHSQLSRQAD	480
DB	350	ERAH-----YSHTMPWRSPSKLS-----HISRLVELSDSDEVDLKSGYMSDSD	394
QY	481	--GKLYALKDNRT-----LQNLSDNKSSEKLVDBIKSYSDVQRQVAILLTDTP	536
DB	395	LMGKTMTEDDITIGWDESSIRSGLSDAONLSSEFNASSLSLPS-----TPTA	447
QY	527	GRHKMSTMPSLDASPESHISLSLHFADAHQGLHGXSELEAQVAISHG-----RLVV	579
DB	448	SRNSTIVLTDSKRLAESGLSWFSESEKAPKLEYDGSGLKMEPEGTSKWRRERPE	507
QY	580	ADSEGKLFSAAP-KQGDGNELKMKAMPQALDEHFGHDHQ-----LSGFPH---DHWGL	631
DB	508	CDSSKGGELKKPISLGHPSGLKKGKTPPVAVTSPTHTHAQSALKVAGKPEGKATDKGL	567
QY	632	NALVRNFRQOHACPLGNDRHFHPGNLTDALVIDNQLGLHNTN-----PRPHEILDM	684
DB	568	--AVKNTGLQRSSSDAGD-----RLSDAKKPPSGIARPSISGSEFYKKPPP-----	612
QY	685	GHLGSLAQBGKLIHYFDOLTGKTGAESDCKQLK--GLDGAAYLLKDGVEKRLINQST	742
DB	613	-ATGTATVMQ-----TGGSATLSKIQKSGSI-----PVKPVNGRKT	648
QY	743	SSIKHGTENVFSLPHVRNKPPEGDALQGLNKDKAQAMAVIG-----VNKYLAL	791
DB	649	LDVNSNAEPCFLAPGARSNIQ-----YRSLPRPAKSSMSVTGGRGPRPVSSSIDPSLLS	704
QY	792	TEKGDIRSFOIKPGT-----QQLERPAQTLRSREGISGELKDHIHVDHKQNLIALTHEGEVF	846
DB	705	TKQGLTPSLKEPTKVASGRITTPAPVNOTDREKAKAKAVALD-SDNI-SLSKIG---	759
QY	847	HQPREAWQNAEASSWHKLA-LFQSRSK-----LKSLDMGHEHKPIATTFDGSQ	894
DB	760	-SPSTPKNAQASHPTATKLAELPPTLRATPAKSFVKPPSLANLDKVNNSLDFPSSDDTT	818

Search completed: July 7, 2004, 15:20:13
 Job time : 95.8129 secs

Search completed: July 7, 2004, 15:20:13
Job time : 95.8129 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:06:21 ; Search time 7.80425 Seconds
(without alignments)
5032.400 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLOYSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	2 AAW98012	AAW98012 Hypersens
2	715	100.0	139	3 AAY71096	Aay71096 Erwinia a
3	715	100.0	139	3 AAY84857	Aay84857 A hyperse
4	715	100.0	139	4 AAU02879	Aau02879 Erwinia a
5	715	100.0	139	5 AAE16450	Aae16450 Erwinia a
6	82	11.5	792	6 ABU20768	Abu20768 Protein e
7	81.5	11.4	402	6 ABU40422	Abu40422 Protein e
8	80.5	11.3	368	2 AAW89180	Aaw89180 Amino aci
9	80.5	11.3	766	3 AAB07739	Aab07739 A snake v
10	80.5	11.3	787	3 AAB07740	Aab07740 A snake v
11	80.5	11.3	820	3 AAB07741	Aab07741 A snake v
12	79.5	11.1	957	7 ADE78250	Ade78250 Endometri
13	79	11.0	789	4 AAG81193	Aag81193 Mycobacte
14	76.5	10.7	1105	2 AAW44864	Aaw44864 Human TPC
15	76.5	10.7	1105	2 AAW73958	Aaw73958 Human TPC
16	76.5	10.7	1105	5 AAE15161	Aae15161 Human TPC
17	76	10.6	1235	5 ABG61674	Abg61674 Cadherin-
18	75.5	10.6	966	5 ABG93047	Abg93047 S. cerevi
19	75	10.5	243	2 AAW38542	Aaw38542 Streptoco
20	75	10.5	317	2 AAY85961	Aay85961 S. pneumo
21	75	10.5	317	3 AAB13516	Aab13516 Streptoco
22	75	10.5	317	4 AAB60873	Aab60873 Mevalonat
23	75	10.5	317	6 ABU00733	Abu00733 S. pneumo
24	75	10.5	317	6 ABU97196	Abu97196 Enzyme po
25	75	10.5	2053	4 ABB71118	Abb71118 Drosophil

26	74.5	10.4	1471	4	ABB58739	Abb58739 Drosophil
27	74	10.3	422	6	ABU41136	Abu41136 Protein e
28	73.5	10.3	984	7	ADE38441	Ad38441 Human pro
29	73.5	10.3	2032	6	ABR43630	Ab43630 Mouse CLA
30	73	10.2	560	5	ABP39498	Abp39498 Staphyloc
31	72.5	10.1	421	6	ABU41975	Abu41975 Protein e
32	72.5	10.1	466	4	ABG63634	Abg63634 Human gas
33	72.5	10.1	590	4	ABG05720	Abg05720 Novel hum
34	72	10.1	584	4	ABG82869	Abg82869 S. epider
35	71.5	10.0	511	7	ADB64185	Ad64185 Human pro
36	71.5	10.0	920	2	AAW82500	Aaw82500 Human OGT
37	71.5	10.0	920	2	ADE80755	Ade80755 Microsate
38	71.5	10.0	920	8	ADE76911	Ade76911 Human pro
39	71.5	10.0	967	6	ABR41060	Ab41060 Human MAP
40	71.5	10.0	1036	6	ABG71215	Abg71215 rat O-lin
41	71.5	10.0	1036	7	ADE62982	Ade62982 Rat Prote
42	71.5	10.0	1591	6	ABR41061	Ab41061 Human MAP
43	70.5	9.9	103	7	ADC95750	Adc95750 E. faeciu
44	70.5	9.9	670	4	AAU04025	Aau04025 Human pro
45	70.5	9.9	988	5	ABG61693	Abg61693 Cadherin-

ALIGNMENTS

RESULT 1

AAW98012
ID AAW98012 standard; protein; 139 AA.

AC AAW98012;

DT 21-JUN-1999 (first entry)

DE Hypersensitive response elicitor DspF.

XX Hypersensitive response elicitor; DspF; disease resistance;

KW insect resistance; biological control; transgenic plant.

XX Erwinia amylovora.

PN WO9907206-A1.

PD 18-FEB-1999.

PF 24-JUL-1998; 98WO-US015426.

PR 06-AUG-1997; 97US-0055105P.

XX (CORR) CORNELL RES FOUND INC.

XX Bogdanove AJ, Kim JF, Wei Z, Beer SV;

XX WPI; 1999-180362/15.

XX N-PSDB; AAX24811.

XX Nucleic acid encoding hypersensitive response-eliciting protein - used to improve growth of plants and impart resistance to disease and insects.

XX Claim 18; Page 57-58; 75pp; English.

XX This polypeptide comprises the 16 kDa hypersensitive response elicitor protein DspF of Erwinia amylovora. The nucleotide sequence of the dsp region of E. amylovora strain Ea321 was determined using subclones of pCPP430. A two-gene operon was discovered comprising dspE (see AAX24810) and dspF (see AAX24811). The isolated dsp DNA molecules and encoded proteins can be used to impart disease resistance to plants, to enhance plant growth, and/or to control insects on plants. This is achieved by applying a hypersensitive response elicitor protein or polypeptide in a non-infectious form to plants or plant seeds, or by producing transgenic plants or plant seeds transformed with DNA encoding a hypersensitive response elicitor. Protection can be provided against a wide range of viruses, bacteria, fungi and insects, e.g. tobacco mosaic virus and tomato mosaic virus, Pseudomonas syringae, Xanthomonas campestris,

CC Fusarium oxysporum, Phytophthora infestans, armyworm, diamondback moth,
CC etc. The method avoids use of infectious agents or polluting chemicals.
CC Claimed transgenic plants are infected from alfalfa, rice, wheat, barley,
CC rye, cotton, sunflower, peanut, corn, potato, bean, pea, chicory,
CC lettuce, endive, cabbage, brussels sprout, sweet potato, beet, parsnip,
CC turnip, cauliflower, broccoli, turnip, radish, spinach, onion, garlic,
CC eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber,
CC apple, pear, melon, citrus, strawberry, grape, raspberry, pineapple,
CC soybean, tobacco, tomato, sorghum, sugarcane, Arabidopsis thaliana,
CC Sainpaulia, petunia, pelargonium, poinsettia, chrysanthemum, carnation
CC and zinnia
XX
SQ Sequence 139 AA;

Query Match 100.0%; Score 715; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.2e-78;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCKRIIE 60
Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCKRIIE 60
QY 61 ADPQTSTILYSMLQLNFEMAAWRCWGLALDELHNVRCLFQOSLEHLDEASFSDIYSGFI 120
Db 61 ADPQTSTILYSMLQLNFEMAAWRCWGLALDELHNVRCLFQOSLEHLDEASFSDIYSGFI 120

QY 121 EHAAEVREYIAQLDESSAA 139
Db 121 EHAAEVREYIAQLDESSAA 139

RESULT 2
AAV71096
ID AAY71096 standard; protein; 139 AA.
XX
AC AAY71096;
DT 08-SEP-2000 (first entry)
XX
DE Erwinia amylovora hypersensitive response elicitor encoded by dspf gene.
XX
KW Hypersensitive response elicitor; environmental stress resistance; plant;
KW dspf gene.
XX
OS Erwinia amylovora.
XX
PN WC200028055-A2.
XX
PD 18-MAY-2000.
XX
XX 04-NOV-1999; 99WO-US026039.
XX
XX 05-NOV-1998; 98US-0107243P.
XX
PA (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI Wei Z, Schading RL;
XX
DR WPI; 2000-376566/32.
DR N-PSDB; AAD00671.
XX
XX Application of a hypersensitive response elicitor protein to plants to
PT impart stress resistance.
PT
PS Disclosure; Page 21; 84pp; English.
XX
XX The patent discloses a method to impart stress resistance to plants by
CC applying a hypersensitive response elicitor in a non-infectious form to a
CC plant or seed. The present sequence is a hypersensitive response elicitor
CC encoded by dspf gene from Erwinia amylovora. The protein is used to
CC impart stress resistance to plants
XX
SQ Sequence 139 AA;

Query Match 100.0%; Score 715; DB 3; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.2e-78;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCKRIIE 60
Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCKRIIE 60
QY 61 ADPQTSTILYSMLQLNFEMAAWRCWGLALDELHNVRCLFQOSLEHLDEASFSDIYSGFI 120
Db 61 ADPQTSTILYSMLQLNFEMAAWRCWGLALDELHNVRCLFQOSLEHLDEASFSDIYSGFI 120
QY 121 EHAAEVREYIAQLDESSAA 139
Db 121 EHAAEVREYIAQLDESSAA 139

RESULT 3
AAY84857
ID AAY84857 standard; protein; 139 AA.
XX
AC AAY84857;
XX
DT 08-AUG-2000 (first entry)
XX
DE A hypersensitive response elicitor protein.
XX
KW Hypersensitive response; insect control; disease resistance;
KW hypersensitive response elicitor; plant growth; vegetable; crop;
KW ornamental plant; dspf gene.
XX
OS Erwinia amylovora.
XX
PN WC200020452-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-US023181.
XX
PR 05-OCT-1998; 98US-0103050P.
XX
PA (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI Wei Z, Fan H, Niggemeyer JL;
XX
DR WPI; 2000-303745/26.
DR N-PSDB; AAA14941.
XX
XX Hypersensitive response elicitor polypeptides useful for imparting
PT enhanced growth, disease resistance and insect resistance to plants,
PT especially vegetables and ornamental flowers.
XX
PS Disclosure; Page 22-23; 100pp; English.

XX
CC The present sequence represents a hypersensitive response elicitor
CC polypeptide. The polynucleotide represents the dspf gene. The
CC specification describes hypersensitive response elicitor polypeptide
CC fragments, which do not elicit a hypersensitive response. Instead, the
CC proteins impart disease resistance to plants, enhance plant growth,
CC and/or control insects. The polypeptide fragments may be used to these
CC properties to plants. The plants which may be treated in this way include
CC vegetables, crops and ornamental plants such as alfalfa, rice, wheat,
CC barley, rye, cotton, sunflower, peanut, corn, potato, sweet potato, bean,
CC pea, chicory, lettuce, endive, cabbage, brussels sprout, beet, parsnip,
CC turnip, cauliflower, broccoli, radish, spinach, zucchini, onion, garlic, eggplant,
CC pepper, celery, carrot, squash, pumpkin, cucumber, apple, pear,
CC melon, citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Sainpaulia, petunia,
CC pelargonium, poinsettia, chrysanthemum, carnation or zinnia
XX
SQ Sequence 139 AA;

Query Match 100.0%; Score 715; DB 3; Length 139;
 Best Local Similarity 100.0%; Pred. No. 4.2e-78;
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
 DB 1 MTSSQQRVERFLQYFSAGCKTPHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60

QY 61 ADPQTSITLYSMLLQLNPFEMAMRGCMWLDLDELHNVRLCFQOSLEHLDEASFSDIVSGFI 120
 DB 61 ADPQTSITLYSMLLQLNPFEMAMRGCMWLDLDELHNVRLCFQOSLEHLDEASFSDIVSGFI 120

QY 121 EHAAEVREYIAQLDESSAA 139
 DB 121 EHAAEVREYIAQLDESSAA 139

RESULT 4
 AAU02879
 ID AAU02879 standard; protein; 139 AA.
 AC AAU02879;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Erwinia amylovora disease-specific region F (dspF) polypeptide.
 XX
 KW Disease-specific region F; dspF; insect control; disease resistance;
 KW hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic;
 KW cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr;
 KW petunia; chrysanthemum; carnation; transgenic plant; avirulence locus;
 KW pathogenicity; fireblight; chaperone.
 XX
 OS Erwinia amylovora.
 XX
 PN USG228644-B1.
 XX
 PD 08-MAY-2001.
 XX
 PF 22-JUL-1998; 98US-00120663.
 XX
 PR 06-AUG-1997; 97US-0055106P.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Bogdanove AJ, Kim JF, Wei Z, Beer SV;
 XX
 DR WPI; 2001-327491/34.
 DR N-PSDB; AAS05200.
 XX
 PT Polynucleotides encoding hypersensitive response eliciting proteins or
 PT polypeptides useful for imparting disease resistance to plants, to
 PT enhance plant growth, and/or to control insects on plants.
 XX
 PS Claim 1; Col 19-20; 37pp; English.
 XX
 CC The sequence represents an Erwinia amylovora disease-specific region F
 CC (dspF) polypeptide, a chaperone of dspE, which elicits a hypersensitive
 CC response in plants. The dspE operon functions as an avirulence (avr)
 CC locus and the dspE protein is required for pathogenicity of Erwinia
 CC amylovora. DspE is also responsible for causing fireblight in certain
 CC plants. The protein and its associated nucleic acid can be applied in a
 CC non-infectious form to plants or plant seeds to impart disease
 CC resistance, to enhance plant growth, and/or to control insects on plants.
 CC Alternatively, plant cells may be transformed with the DNA to form
 CC transgenic plants with the same properties. The method can be utilised to
 CC treat a wide variety of plants and seeds, including crop plants such as
 CC rice, wheat, corn, cabbage, cauliflower, garlic, cucumber, apple, grape,
 CC tobacco and sugarcane, and ornamental plants such as Arabidopsis
 CC thaliana, petunia, chrysanthemum and carnation
 XX
 SQ Sequence 139 AA;

Query Match 100.0%; Score 715; DB 4; Length 139;
 Best Local Similarity 100.0%; Pred. No. 4.2e-78;
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
 DB 1 MTSSQQRVERFLQYFSAGCKTPHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60

QY 61 ADPQTSITLYSMLLQLNPFEMAMRGCMWLDLDELHNVRLCFQOSLEHLDEASFSDIVSGFI 120
 DB 61 ADPQTSITLYSMLLQLNPFEMAMRGCMWLDLDELHNVRLCFQOSLEHLDEASFSDIVSGFI 120

QY 121 EHAAEVREYIAQLDESSAA 139
 DB 121 EHAAEVREYIAQLDESSAA 139

RESULT 5
 AAU16450
 ID AAU16450 standard; protein; 139 AA.
 AC AAU16450;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Erwinia amylovora hypersensitive response elicitor protein, dspF.
 XX
 KW Hypersensitive response elicitor protein; plant growth; fruit coloration;
 KW disease resistance; stress resistance; phytotoxin; insect infection;
 KW plant maturation; dspF protein.
 XX
 OS Erwinia amylovora.
 XX
 PN WO200198501-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 12-JUN-2001; 2001WO-US018820.
 XX
 PR 16-JUN-2000; 2000US-0212211P.
 XX
 PA (EDEN-) EDEN BIOSCIENCE CORP.
 XX
 PI Fan H, Wei Z;
 XX
 DR WPI; 2002-122282/16.
 DR N-PSDB; AAD27018.
 XX
 PT New hypersensitive response elicitor proteins comprising spaced apart
 PT domains having an acidic portion linked to an alpha-helix, useful for
 PT imparting disease or stress resistance, controlling insects or enhancing
 PT plant growth.
 XX
 PS Disclosure; Page 24; 99pp; English.
 XX
 CC The patent discloses hypersensitive response elicitor proteins and
 CC nucleotides encoding such proteins. Hypersensitive response elicitor
 CC proteins comprise an isolated pair or more of spaced apart domains, each
 CC comprising an acidic portion linked to an alpha-helix and capable of
 CC eliciting a hypersensitive response in plants. Sequences of the invention
 CC are used to impart disease resistance to plants, to enhance plant growth,
 CC to control insects and/or to impart stress resistance to plants which
 CC includes resistance to environmental stresses such as climate, air
 CC pollution, chemical and nutritional stress. The method of imparting
 CC disease resistance has the potential for treating previously untreatable
 CC diseases, treating diseases systemically and avoiding the use of
 CC infectious agents or environmentally harmful materials. Hyper- sensitive
 CC response elicitor sequences are used to enhance plant growth which
 CC encompasses greater yield, increased in quantity of seeds produced,
 CC percentage of seeds germinated, plant size and biomass, bigger fruits,
 CC earlier fruit coloration and plant maturation. They are also used for
 CC insect control which encompasses preventing direct insect damage to plant
 CC by feeding injury, interfering with insect larval feeding on the plants,

CC preventing insects from colonising host plants and releasing phytotoxins.
 CC Sequences of the invention also prevent subsequent disease damage to
 CC plants resulting from insect infection. The present sequence is *Erwinia*
 CC amylovora hypersensitive response elicitor protein, dspF

XX Sequence 139 AA;

Query Match 100.0%; Score 715; DB 5; Length 139;
 Best Local Similarity 100.0%; Pred. No. 4.2e-78;
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTSSQORVERFLOYSAGCKTPHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRHIE 60
 Db 1 MTSSQORVERFLOYSAGCKTPHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRHIE 60
 QY 61 ADPQTSITLYSMLLQNFENMAARGCWLDELHNVRLCFQQLSLHLEASFSDIYSGFI 120
 Db 61 ADPQTSITLYSMLLQNFENMAARGCWLDELHNVRLCFQQLSLHLEASFSDIYSGFI 120
 QY 121 EHAAVREYIAQDESSAA 139
 Db 121 EHAAVREYIAQDESSAA 139

RESULT 6

ABU20768
 ID ABU20768 standard; protein; 792 AA.

XX AC ABU20768;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #6295.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Bacteroides fragilis.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA24638.

XX XX New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids, required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX XX Claim 25; SEQ ID NO 48692; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 792 AA;

Query Match 11.5%; Score 82; DB 6; Length 792;

Best Local Similarity 22.5%; Pred. No. 2;

Matches 42; Conservative 27; Mismatches 42; Indels 76; Gaps 10;

QY 1 MTSSQORVER-----FLQFSAGCKTPHLKDGVCALYN----- 34

Db 535 MQNIQSVREHOLIVAKQMAEKALQKQSLNNMSHEIRTPL---NAIVGFTNVLGEGS 591

QY 35 ---EQDEAAVLEVPQHSLSLLHCRHIEADPQTSITLYSMLLQ-----LNFEVMAAM 83

Db 592 EIDPDEKASMLEIHNHNLLE-----KLNDVLEIRLSDSGSLDFDMKE- 637

QY 84 RGCWLALD---ELHN-----VRLCFQQLSLE-----HLDEAFSFDIVSGFIEHAAEV 126

Db 638 ---WNMTDIVKEIYKTYQPLIRLSLQFRLELDDTVSVVPHTRLRFVQVISFLNNANKF 694

QY 127 RE--YIA 131

Db 695 TQNGYIA 701

RESULT 7

ABU40422

ID ABU40422 standard; protein; 402 AA.

XX AC ABU40422;

XX XX 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #25949.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Pseudomonas putida.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

```
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX
XX
XX WPI: 2003-029926/02.
XX N-PSDB; ACA44292.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 68346; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 402 AA;
XX
XX Query Match 11.4%; Score 81.5; DB 6; Length 402;
XX Best Local Similarity 20.5%; Pred. No. 0.88;
XX Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;
XX
XX QY 12 LQYFAGCKTPI--HLKDGVCALYNEQDEAAVLEVPQHSQ----- 50
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 192 IERISRDARVPVKHL-DGICHY-----VQHALDKAMVAVNAKTYRGIC 239
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX QY 51 ----SLLLHCRILPADPQTSTLTYSMLQLNFEMAMRG-----MLA- 89
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 240 GAMETLLVDQQAER-----FLPEMARRFVEKGVELRGCEQTQAIIISAKPATADWHT 293
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX QY 90 -LDELHNVRLC--FQSLHLEDEASFSQDVGFTIEHAAVEVREYIAQLDESS 137
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 294 YLDAILSIRVVYDGLNQLAIEHINHYGSHHTDSIISEHQGEARQFMAEVDSDAS 344
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 8
XX AAW89180
XX ID AAW89180 standard; protein; 368 AA.
XX
XX AC AAW89180;
XX
XX DT 11-MAR-1999 (first entry)
XX
XX DE Amino acid sequence of delta9 14:0-ACP desaturase.
```

```
XX
XX delta 9 14:0-ACP desaturase; fatty acid; anacardic acid; pest;
XX resistance; plant; pelargonium; geranium; polymeric; nylon.
XX
XX OS Pelargonium sp.
XX
XX PN US856157-A.
XX
XX PD 05-JAN-1999.
XX
XX PF 04-JUN-1997; 97US-00869137.
XX
XX PR 04-JUN-1996; 96US-0018957P.
XX
XX PA (PENN-) PENN STATE RES FOUND.
XX
XX PI Cox-Foster DL, Craig R, Schultz D, Medford JI, Mumma RO;
XX N-PSDB; AAV81284.
XX
XX WPI: 1999-105114/09.
XX N-PSDB; AAV81284.
XX
XX New isolated 14:0-ACP desaturase gene - obtained from geranium plants,
XX used to enhance pest resistance of plants and for enhancing production of
XX unsaturated fatty acids in plants.
XX
XX Claim 1; Fig 3; 26pp; English.
XX
XX This represents a delta9 14:0-ACP desaturase. The 14:0-ACP desaturase
XX gene enhances the accumulation 16:1 Delta-11 and 18:1 Delta'3 fatty acids
XX and 22:1 omega 5 and 24:1 omega 5 anacardic acids. The gene can be used
XX to provide pest resistance in plants. The gene can also be used to
XX enhance the production of unsaturated fatty acids in plants such as
XX soybeans, rapeseed, maize, sunflower, safflower, cotton, cuphea, peanut,
XX coconut, oil-palm and pelargonium. The gene and its expression products
XX can also be used in manufacturing methods, e.g. in innovative processes
XX for making specialty polymers such as nylon and other methods in which
XX unsaturated fatty acids are used as constituents or starting materials.
XX In addition the promoter from the gene can be used for trichome specific
XX expression of polypeptides
XX
XX SQ Sequence 368 AA;
XX
XX Query Match 11.3%; Score 80.5; DB 2; Length 368;
XX Best Local Similarity 20.1%; Pred. No. 1;
XX Matches 37; Conservative 26; Mismatches 54; Indels 67; Gaps 8;
XX
XX QY 6 QRVERFLOYFSA-----GC-KTPIHLKDGVCALYNEQDEAAVLE-----VFOHSDSL 52
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 169 RQVEKTIQYLIALGQDIGTEKNPYHL-----FIYTSFQERATFISHANTAKLAQHQGDKQ 223
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX QY 53 LLH-CRIIEADPQTSITLYSMLQLNFEM----- 80
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 224 LAQICGTIADEKREHETAYTRIVDKLFELDPDETMSCLAHMMKKITWPAHLMRDGRDPH 283
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX QY 81 -----AAMRGCWALDELHNVRLCFQOSLEH-LDEASFSQDVGFTIEHAAVEVREYIA 131
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 284 LFQHFVSVASRTGYVTWMDVIN-----ILEHFVEKWNIEKITAGLSKDGREAQDYVC 335
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX QY 132 QLDE 135
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 336 KLGE 339
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 9
XX AAB07739
XX ID AAB07739 standard; protein; 766 AA.
XX
XX AC AAB07739;
XX
XX DT 07-NOV-2000 (first entry)
XX
XX DE A snake venom protease (SVPH-1) polypeptide variant SVPH-la.
XX
```

KW Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin; chromosome 1; chromosome 4; immune system; splice variant.
 XX
 XX
 OS Homo sapiens.
 XX WO200043525-A2.
 XX 27-JUL-2000.
 PD
 XX 21-JAN-2000; 2000WO-US001338.
 XX
 XX 21-JAN-1999; 99US-0116670P.
 PR 14-JUN-1999; 99US-0138682P.
 PR 27-SEP-1999; 99US-0155798P.
 XX
 XX (IMMV) IMMUNEX CORP.
 PA
 XX Cerretti DP;
 PI
 XX WPI; 2000-482914/42.
 DR N-PSDB; AAAS9304.
 XX
 XX Snake venom protease (SVPH) nucleic acids, and polypeptides, used to identify proteins having metalloproteinase-disintegrin activity, and inhibitors of the proteins for use in therapeutics.
 PT
 PT
 PT
 XX Claim 13; Page 15; 105pp; English.
 XX
 XX The present sequence represents an alternatively spliced snake venom protease-1 (SVPH-1) polypeptide, designated SVPH-1a. The SVPH polypeptides are metalloproteinase-disintegrin protein family members. The SVPH polynucleotides can be used as probes to identify nucleic acids encoding proteins having metalloproteinase-disintegrin activity, to identify human chromosome 1 or 4, to map genes on those chromosomes, to identify genes associated with diseases, syndromes and conditions associated with the chromosomes, and to study proteinases and their activities on cell/cell interactions and the immune system. Sense or antisense oligonucleotides of SVPH can be used to inhibit gene expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cell/cell and cell/matrix interactions involved in cellular processes and in the immune system. The polypeptides may also be used to screen for inhibitors of the polypeptide's activity, which are used in therapeutics. The antibodies can be used in assays to detect the presence of the polypeptides in vitro or in vivo, and to purify the polypeptides by affinity chromatography
 XX
 SQ Sequence 766 AA;
 Query Match 11.3%; Score 80.5; DB 3; Length 766;
 Best Local Similarity 22.3%; Pred. No. 2.9;
 Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;
 QY 18 GCKTPHLK-----DGVCALYNEQDEAAVLEVPQHSLSLLHCRIRIADPQTSITL-- 69
 DB 60 GQKHITHIKVKLLFSKHLFPVFTYDQGAILEDQPFVQNNCYH-GYVEGDPESLSLST 118
 QY 70 ----YSMLLQLNFEMAMRGCLALDELHNVR-LCFQOSLEHL-----DEAGSFSDIVSG 118
 DB 119 CFGGFGQILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTWRSG 165
 QY 119 FIEHAAEVREYIAQLDESS 137
 DB 166 FMQNEITCRMEFEIDNST 184
 RESULT 10
 AAB07740
 ID AAB07740 standard; protein; 787 AA.
 XX
 XX AAB07740;
 XX
 XX 07-NOV-2000 (first entry)
 XX

DE A snake venom protease (SVPH-1) polypeptide variant SVPH-1b.
 XX
 KW Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin; chromosome 1; chromosome 4; immune system; splice variant.
 XX
 OS Homo sapiens.
 XX WO200043525-A2.
 XX 27-JUL-2000.
 PD
 XX 21-JAN-2000; 2000WO-US001338.
 XX
 XX 21-JAN-1999; 99US-0116670P.
 PR 14-JUN-1999; 99US-0138682P.
 PR 27-SEP-1999; 99US-0155798P.
 XX
 XX (IMMV) IMMUNEX CORP.
 PA
 XX Cerretti DP;
 PI
 XX WPI; 2000-482914/42.
 DR N-PSDB; AAAS9305.
 XX
 XX Snake venom protease (SVPH) nucleic acids, and polypeptides, used to identify proteins having metalloproteinase-disintegrin activity, and inhibitors of the proteins for use in therapeutics.
 PT
 PT
 PT
 XX Claim 13; Page 15-16; 105pp; English.
 XX
 XX The present sequence represents an alternatively spliced snake venom protease-1 (SVPH-1) polypeptide, designated SVPH-1b. The SVPH polypeptides are metalloproteinase-disintegrin protein family members. The SVPH polynucleotides can be used as probes to identify nucleic acids encoding proteins having metalloproteinase-disintegrin activity, to identify human chromosome 1 or 4, to map genes on those chromosomes, to identify genes associated with diseases, syndromes and conditions associated with the chromosomes, and to study proteinases and their activities on cell/cell interactions and the immune system. Sense or antisense oligonucleotides of SVPH can be used to inhibit gene expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cell/cell and cell/matrix interactions involved in cellular processes and in the immune system. The polypeptides may also be used to screen for inhibitors of the polypeptide's activity, which are used in therapeutics. The antibodies can be used in assays to detect the presence of the polypeptides in vitro or in vivo, and to purify the polypeptides by affinity chromatography
 XX
 SQ Sequence 787 AA;
 Query Match 11.3%; Score 80.5; DB 3; Length 787;
 Best Local Similarity 22.3%; Pred. No. 3;
 Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;
 QY 18 GCKTPHLK-----DGVCALYNEQDEAAVLEVPQHSLSLLHCRIRIADPQTSITL-- 69
 DB 60 GQKHITHIKVKLLFSKHLFPVFTYDQGAILEDQPFVQNNCYH-GYVEGDPESLSLST 118
 QY 70 ----YSMLLQLNFEMAMRGCLALDELHNVR-LCFQOSLEHL-----DEAGSFSDIVSG 118
 DB 119 CFGGFGQILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTWRSG 165
 QY 119 FIEHAAEVREYIAQLDESS 137
 DB 166 FMQNEITCRMEFEIDNST 184
 RESULT 11
 AAB07741
 ID AAB07741 standard; protein; 820 AA.
 XX
 XX AAB07741;
 XX

DT 07-NOV-2000 (first entry)
XX
DE A snake venom protease (SVPH-1) polypeptide variant SVPH-1c.
XX
KW Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;
XX chromosome 1; chromosome 4; immune system; splice variant.
XX
OS Homo sapiens.
XX
XX W0200043525-A2.
XX
PD 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-US001338.
XX
XX 21-JAN-1999; 99US-0116670P.
PR 14-JUN-1999; 99US-0138682P.
PR 27-SEP-1999; 99US-0135798P.
XX
XX (IMMV) IMMUNEX CORP.
XX
XX Cerretti DP;
XX
XX WPI; 2000-482914/42.
DR N-PSDB; AAA59306.
XX
XX Snake venom protease (SVPH) nucleic acids, and polypeptides, used to
PT identify proteins having metalloproteinase-disintegrin activity, and
PT inhibitors of the proteins for use in therapeutics.
XX
PS Claim 13; Page 16; 105pp; English.
XX
XX The present sequence represents an alternatively spliced snake venom
CC protease-1 (SVPH-1) polypeptide, designated SVPH-1c. The SVPH
CC polypeptides are metalloproteinase-disintegrin protein family members.
CC The SVPH polynucleotides can be used as probes to identify nucleic acids
CC encoding proteins having metalloproteinase-disintegrin activity, to
CC identify human chromosome 1 or 4, to map genes on those chromosomes, to
CC identify genes associated with diseases, syndromes and conditions
CC associated with the chromosomes, and to study proteinases and their
CC activities on cell/cell interactions and the immune system. Sense or
CC antisense oligonucleotides of SVPH can be used to inhibit gene expression
CC of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cell/cell
CC and cell/matrix interactions involved in cellular processes and in the
CC immune system. The polypeptides may also be used to screen for inhibitors
CC of the polypeptide's activity, which are used in therapeutics. The
CC antibodies can be used in assays to detect the presence of the
CC polypeptides in vitro or in vivo, and to purify the polypeptides by
CC affinity chromatography
XX
SQ Sequence 820 AA;

Query Match 11.3%; Score 80.5; DB 3; Length 820;
Best Local Similarity 22.3%; Pred. No. 3.2;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPIHLK-----DGVCALYNEQDERAAVLEVPQHSDDLHCRIFEADPQTSITL-- 69
DB 60 GQKHIIHKVKKLFESKHLVFVTYDQGAILEDQPFVQNCYYH-GYVEGDPESILVST 118

QY 70 ----YSMLLQINFMAAMRGCGWLALDELHNVRLCFQOSLEHLDEASFSDIVSGFIEHAAE 118
DB 119 CFGGFGQILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTRSG 165

QY 119 FIEHAAEVREYIAQLDESS 137
DB 166 FMQNEITCRMEFEIDNST 184

RESULT 12
ADE78250
ID ADE78250 standard; protein; 957 AA.
XX

AC ADE78250;
XX
DT 29-JAN-2004 (first entry)
XX
DE Endometrial specific protein identified as DEX0379_11_aa_15 (SeqID 157).
XX
KW neoplastic; endometrial cell; endometrial specific nucleic acid; ESNA;
XX cancer; metastasis; gene therapy; cytostatic; human.
XX
OS Homo sapiens.
XX
XX W02003055982-A2.
XX
PD 10-JUL-2003.
XX
XX 20-DEC-2002; 2002WO-US041175.
XX
XX 21-DEC-2001; 2001US-0343134P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Sun Y, Liu C;
XX
XX WPI; 2003-569441/53.
DR N-PSDB; ADE78347.
XX
XX New nucleic acid, useful for preparing a composition for diagnosing or
PT treating endometrial cancer.
PT
XX
XX Claim 1; SEQ ID NO 157; 327pp; English.
XX
XX This invention relates to novel nucleic acid molecules, and encoded
CC proteins thereof, which are present in normal and neoplastic endometrial
CC cells. Specifically, it refers to the use of these endometrial specific
CC nucleic acids (ESNAs), as well as suitable antibodies, agonists and
CC antagonists that are useful for the identification, diagnosis and
CC monitoring of endometrial cancer. The present invention describes a
CC method for monitoring the presence of an endometrial specific protein in
CC the sample with a reagent and detecting the intensity of this
CC interaction. Accordingly, via gene therapy, these ESNAs are useful for
CC preparing a cytostatic composition for diagnosing or treating endometrial
CC cancer. This polypeptide sequence is a human endometrial specific
CC polypeptide of the invention.
XX
SQ Sequence 957 AA;

Query Match 11.1%; Score 79.5; DB 7; Length 957;
Best Local Similarity 32.8%; Pred. No. 5.2;
Matches 21; Conservative 12; Mismatches 16; Indels 15; Gaps 3;

QY 67 ITLYSMILLQINFMAAMRGCGWLALDELHNVRLCFQOSLEHLDEASFSDIVSGFIEHAAE 125
DB 354 ILFPAICLVLPFET-----ILVNEIYNYVLICFQ-----EAPFDVANGFIEEIIH 399

QY 126 VREY 129
DB 400 FKNY 403

RESULT 13
AAG81193
ID AAG81193 standard; protein; 789 AA.
XX
XX AAG81193;
XX
XX 04-SEP-2001 (first entry)
XX
XX Mycobacterium tuberculosis potential drug target protein SEQ ID 244.
XX
XX Drug target; growth; organism viability; characterisation.
XX
XX Mycobacterium tuberculosis.
XX

XX WO200135317-A1.
 XX 17-MAY-2001.
 XX
 XX 13-NOV-2000; 2000WO-US031152.
 XX
 XX 12-NOV-1999; 99US-0165086P.
 PR 12-NOV-1999; 99US-0165124P.
 PR 01-FEB-2000; 2000US-0179531P.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Elsenberg D, Rotstein SH, Marcotte EM;
 XX
 XX WPI; 2001-329193/34.
 DR N-PSDB; AAH52044.
 XX
 XX Identifying nucleotide or polypeptide sequence for use as drug target,
 PT involves providing algorithm that analyzes a functional relationship
 PT between nucleotide or polypeptide sequences, and comparing the sequences.
 XX
 XX Disclosure; Page 178; 207pp; English.
 XX
 XX This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism
 XX
 XX Sequence 789 AA;
 SQ
 Query Match 11.0%; Score 79; DB 4; Length 789;
 Best Local Similarity 24.3%; Fred. No. 4.6;
 Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;
 QY 23 IHLKDGVCALYNEQDEAAVLEVPQSDSLLLHCRHIEADPQTSITLYSMILQINFEMAA 82
 Db 445 VSMRQYLGPHGLTQDPKRLALQKMSFEVAVRILQATPVATGLVSAALL-----LT 498
 QY 83 MRGCVLALDELHNVRLCFQOSLEHLD-----EASFSDIVSGF---- 119
 Db 499 TRGTALTLDQLHHT--LQSDLDYLRKQSPVSTSAFLRLRSREGVRAAADALSGNHPVTR 555
 QY 120 IEHAAVREYIAQLDESSAA 139
 Db 556 VDSGRFPVWYIAPDDEHAAA 575
 RESULT 14
 AA44864
 ID AA44864 standard; protein; 1105 AA.
 XX
 XX AA44864;
 AC
 XX
 XX 28-AUG-1998 (first entry)
 XX
 XX Human TPC2 telomere length and telomerase regulatory protein.
 XX
 XX TPC2; telomere length; telomerase; human; cancer; gene therapy;
 KW diagnosis; vaccine.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH

FT Misc-difference 182 /note= "encoded by TTT"
 FT Misc-difference 211 /note= "encoded by NGG"
 FT
 FT Misc-difference 217 /note= "encoded by GCT, C being uncertain"
 FT Misc-difference 231 /note= "encoded by ATC, C being uncertain"
 FT Misc-difference 381 /note= "encoded by TTA, T being uncertain"
 FT Misc-difference 383 /note= "encoded by AYT"
 FT Misc-difference 387 /note= "encoded by GGC, the second G being uncertain"
 FT Misc-difference 396 /note= "encoded by GAG, the second G being uncertain"
 FT Misc-difference 403 /note= "encoded by CYT"
 FT Misc-difference 404 /note= "encoded by TCA, C being uncertain"
 FT Misc-difference 409 /note= "encoded by GAA, G being uncertain"
 FT Misc-difference 410 /note= "encoded by CAG, G being uncertain"
 FT Misc-difference 413 /note= "encoded by ACT, C being uncertain"
 FT Misc-difference 416 /note= "encoded by TAC, A being uncertain"
 FT Misc-difference 419 /note= "encoded by GCT, C being uncertain"
 FT Misc-difference 421 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 426 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 434 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 439 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 449 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 450 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 454 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 464 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 469 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 476 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 477 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 487 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 488 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 491 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 510 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 515 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 517 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 527 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 529 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 530 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 551 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 562 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 563 /note= "encoded by GCT, the second C being uncertain"
 FT Misc-difference 569 /note= "encoded by GCT, the second C being uncertain"

Search completed: July 7, 2004, 15:17:56
Job time : 9.80425 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:15:47 ; Search time 2.67172 Seconds
(without alignments)
2685.909 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....TEHAAEVREXIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	3	US-09-120-663-4
2	715	100.0	139	4	US-09-431-614-10
3	80.5	11.3	368	2	US-08-869-137-2
4	76.5	10.7	1105	2	US-08-710-249-2
5	76.5	10.7	1105	4	US-09-220-157A-2
6	75	10.5	243	4	US-08-858-207A-338
7	74	10.3	424	4	US-09-543-681A-7210
8	74	10.3	589	4	US-09-540-236-3398
9	73.5	10.3	984	2	US-08-673-789-9
10	73.5	10.3	984	2	US-08-449-645A-19
11	73.5	10.3	984	2	US-08-702-367A-19
12	73.5	10.3	984	5	PCT-US95-04681-19
13	73	10.2	217	4	US-09-489-039A-7789
14	73	10.2	560	4	US-09-134-001C-4343
15	71.5	10.0	572	4	US-09-543-681A-8138
16	70.5	9.9	103	4	US-09-107-532A-5377
17	70	9.8	946	4	US-09-657-931A-10
18	69.5	9.7	2183	1	US-08-348-891A-7
19	69.5	9.7	2183	2	US-09-905-817-7
20	68.5	9.6	107	4	US-09-489-039A-9798
21	68.5	9.6	239	4	US-09-485-885-12
22	68.5	9.6	591	1	US-08-179-738-5
23	68.5	9.6	591	1	US-08-179-738-10
24	68.5	9.6	591	2	US-08-628-145-5
25	68.5	9.6	591	2	US-08-628-145-10
26	68.5	9.6	595	1	US-08-171-718-16
27	68.5	9.6	595	3	US-08-478-087-16

28 68.5 9.6 596 1 US-08-179-738-2 Sequence 2, Appli
29 68.5 9.6 596 1 US-08-179-738-3 Sequence 3, Appli
30 68.5 9.6 596 2 US-08-628-145-2 Sequence 2, Appli
31 68.5 9.6 596 2 US-08-628-145-3 Sequence 3, Appli
32 68 9.5 225 4 US-09-198-452A-80 Sequence 80, Appli
33 68 9.5 371 4 US-09-543-681A-7332 Sequence 7332, Ap
34 68 9.5 652 1 US-08-261-663A-6 Sequence 6, Appli
35 68 9.5 652 4 US-09-357-206A-5 Sequence 5, Appli
36 68 9.5 652 4 US-09-813-742A-5 Sequence 5, Appli
37 68 9.5 652 4 PCT-US95-07754A-6 Sequence 6, Appli
38 68 9.5 1143 2 US-08-310-912A-108 Sequence 108, App
39 68 9.5 1143 3 US-09-301-085-108 Sequence 108, App
40 68 9.5 1143 5 PCT-US95-04589-108 Sequence 108, App
41 68 9.5 1144 1 US-08-261-663A-2 Sequence 2, Appli
42 68 9.5 1144 1 US-08-261-663A-4 Sequence 4, Appli
43 68 9.5 1144 3 US-08-930-996A-9 Sequence 9, Appli
44 68 9.5 1144 4 US-09-357-206A-3 Sequence 3, Appli
45 68 9.5 1144 4 US-09-813-742A-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-120-663-4
; Sequence 4, Application US/09120663
; Patent No. 6228644
; GENERAL INFORMATION:
; APPLICANT: Bogdanove, Adam J.
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/120,663
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,105
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1661
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 139 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-120-663-4

Query Match 100.0%; Score 715; DB 3; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.2e-81;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

RESULT 5
US-09-220-157A-2
; Sequence 2, Application US/09220157A
; Patent No. 6300110
; GENERAL INFORMATION:
; APPLICANT: Villeponteau, Bryant
; APPLICANT: Feng, Junli
; APPLICANT: Andrews, William H.
; APPLICANT: Adams, Robert R.
; TITLE OF INVENTION: Methods and Reagents for Regulating
; TITLE OF INVENTION: Telomere Length and Telomerase Activity
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

RESULT 6
US-08-858-207A-338
; Sequence 338, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```



```

Db      94 RVHVELQFTVRDCKSPFGGAGPLGCKETFNLLYMSDQDVGI-----QRRPLFQKVTTVA 149
QY      61 ADPQTSI-TLYSMLLQLNFEMAAM-----RCGWALDE-----LHNVRLCFQQSLEHLD 108
Db      150 ADQSFTRDLASGSGVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVYQRCPETLN 209
QY      109 E-ASFSDIVSG 118
Db      210 GLAQFPDTLPG 220

RESULT 11
US-08-702-367A-19
; Sequence 19, Application US/08702367A
; Patent No. 5981246
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,367A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-702-367A-19

Query Match 10.3%; Score 73.5; DB 2; Length 984;
Best Local Similarity 27.5%; Pred. No. 4.4;
Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6

QY      7 RVRRPFIQYFSAGKT-----PVLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRILIE 60
Db      94 RVHVELQFTVRDCKSPFGGAGPLGCKETFNLLYMSDQDVGI-----QRRPLFQKVTTVA 149
QY      61 ADPQTSI-TLYSMLLQLNFEMAAM-----RCGWALDE-----LHNVRLCFQQSLEHLD 108
Db      150 ADQSFTRDLASGSGVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVYQRCPETLN 209
QY      109 E-ASFSDIVSG 118
Db      210 GLAQFPDTLPG 220

RESULT 12
PCT-US95-04681-19
; Sequence 19, Application PC/TUS9504681
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 28

```


; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amgen Patent Operations/RBW
 ; STREET: 1840 Behavilland Drive
 ; CITY: Thousand Oaks
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 91320
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04681
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Winter, Robert B.
 ; REFERENCE/DOCKET NUMBER: A-287
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 984 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-04681-19
 ;
 ; Query Match 10.3%; Score 73.5; DB 5; Length 984;
 ; Best Local Similarity 27.5%; Pred. No. 4.4; Indels 23; Gaps 6;
 ; Matches 36; Conservative 20; Mismatches 52;
 ;
 ; 7 RVERFLQYFSGCKT-----PIHLKGVCAALYNEQDEEAAVLEVPQHSLLHLCRIIE 60
 ; 94 RVHVELQFTVRDCKSPFGAGPCKCKETFNLLYMSDQDVGI-----QLRPLFKQVTTVA 149
 ; 61 ADPQTSI-TLYSMLLQINFEMAAM-----RGCWIALDE-----LHNVRLCFQOQSLEHLD 108
 ; 150 ADQSFTIRDLASGVKLVNRCSLGRLTRRGLVLAHPNPGACVALSVRVFYQRCPTLN 209
 ; 109 E-ASFSDIVSG 118
 ; 210 GLAQFFDTLEG 220
 ;
 ; RESULT 13
 ; US-09-489-039A-7789
 ; Sequence 7789, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 7789
 ; LENGTH: 217
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-7789
 ;
 ; Query Match 10.2%; Score 73; DB 4; Length 217;
 ; Best Local Similarity 22.1%; Pred. No. 0.58; Indels 37; Gaps 5;
 ; Matches 33; Conservative 17; Mismatches 37;
 ;
 ; 28 GVCALYNEQDEEAAVLEVPQ-----HSDSLILHC-----RIIEADPQTSITILYS- 71
 ; 2 GVCVAKNYDITLALAGVCQARLVQLAHQCHCDSDALVLSNIIIDLPSTLAVFGG 61
 ;
 ; 72 -----MLLQLNFEMAAEGCGWLALDELHNVRL 98
 ; 62 SEANLRGLGLETLLGVNTSSRQGLNAELRYTILSNVLERKLAASKG---AMDTLGNRIA 118
 ; 99 CFQOSLEHLDEAS-----FSDIVS 117
 ; 119 GLHRQLEHFDLQSETLLSNMAGIYVDVIS 147
 ;
 ; RESULT 14
 ; US-09-134-001C-4343
 ; Sequence 4343, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4343
 ; LENGTH: 560
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4343
 ;
 ; Query Match 10.2%; Score 73; DB 4; Length 560;
 ; Best Local Similarity 26.7%; Pred. No. 2.3; Indels 30; Gaps 6;
 ; Matches 31; Conservative 19; Mismatches 36;
 ;
 ; 34 NEQDEEA--AVLEVPQHSLLL-----HCRITIEADPQTSITILYS-----MLLQLNFE 79
 ; 276 NPEDHKAFQOSIELANHIDADLLIGTDPADRIGIVERDAEAGNIHYNGNQIGALLNLR 335
 ; 80 MAAMRGCGWLALDELHNVRLCFQ-----QSLHLEDEASFSDIVSGFIEHAAEVR 127
 ; 336 IKQTEG-----LSN-RIMFQISVSGGLAKSLAQYHNVNFKVLTGFKYIAAEIR 383
 ;
 ; RESULT 15
 ; US-09-543-681A-8138
 ; Sequence 8138, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 8138
 ; LENGTH: 572
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-8138
 ;
 ; Query Match 10.0%; Score 71.5; DB 4; Length 572;
 ; Best Local Similarity 20.9%; Pred. No. 3.6; Indels 58; Gaps 7;
 ; Matches 41; Conservative 32; Mismatches 58;
 ;
 ; 4 SQORVE----RFLQYFSGCKTPIHLKDGVCALYNEQ-----DEEAAVLEVPQ 48
 ; 125 SQEIRNLKGRFIEYFSAALNQLLEILLQDKNLFQEFQFTTSYQDGFYKYLAYYMRNEKF 184
 ; 49 SDSLLHLCRIIEADPQTSITILYSMLLQLNFEMAAEGCGWLALD----- 91

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amgen Patent Operations/RBW
 ; STREET: 1840 Behavilland Drive
 ; CITY: Thousand Oaks
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 91320
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04681
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Winter, Robert B.
 ; REFERENCE/DOCKET NUMBER: A-287
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 984 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-04681-19
 ;
 ; Query Match 10.3%; Score 73.5; DB 5; Length 984;
 ; Best Local Similarity 27.5%; Pred. No. 4.4; Indels 23; Gaps 6;
 ; Matches 36; Conservative 20; Mismatches 52;
 ;
 ; 7 RVERFLQYFSGCKT-----PIHLKGVCAALYNEQDEEAAVLEVPQHSLLHLCRIIE 60
 ; 94 RVHVELQFTVRDCKSPFGAGPCKCKETFNLLYMSDQDVGI-----QLRPLFKQVTTVA 149
 ; 61 ADPQTSI-TLYSMLLQINFEMAAM-----RGCWIALDE-----LHNVRLCFQOQSLEHLD 108
 ; 150 ADQSFTIRDLASGVKLVNRCSLGRLTRRGLVLAHPNPGACVALSVRVFYQRCPTLN 209
 ; 109 E-ASFSDIVSG 118
 ; 210 GLAQFFDTLEG 220
 ;
 ; RESULT 13
 ; US-09-489-039A-7789
 ; Sequence 7789, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 7789
 ; LENGTH: 217
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-7789
 ;
 ; Query Match 10.2%; Score 73; DB 4; Length 217;
 ; Best Local Similarity 22.1%; Pred. No. 0.58; Indels 37; Gaps 5;
 ; Matches 33; Conservative 17; Mismatches 37;
 ;
 ; 28 GVCALYNEQDEEAAVLEVPQ-----HSDSLILHC-----RIIEADPQTSITILYS- 71
 ; 2 GVCVAKNYDITLALAGVCQARLVQLAHQCHCDSDALVLSNIIIDLPSTLAVFGG 61
 ;
 ; 72 -----MLLQLNFEMAAEGCGWLALDELHNVRL 98
 ; 62 SEANLRGLGLETLLGVNTSSRQGLNAELRYTILSNVLERKLAASKG---AMDTLGNRIA 118
 ; 99 CFQOSLEHLDEAS-----FSDIVS 117
 ; 119 GLHRQLEHFDLQSETLLSNMAGIYVDVIS 147
 ;
 ; RESULT 14
 ; US-09-134-001C-4343
 ; Sequence 4343, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4343
 ; LENGTH: 560
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4343
 ;
 ; Query Match 10.2%; Score 73; DB 4; Length 560;
 ; Best Local Similarity 26.7%; Pred. No. 2.3; Indels 30; Gaps 6;
 ; Matches 31; Conservative 19; Mismatches 36;
 ;
 ; 34 NEQDEEA--AVLEVPQHSLLL-----HCRITIEADPQTSITILYS-----MLLQLNFE 79
 ; 276 NPEDHKAFQOSIELANHIDADLLIGTDPADRIGIVERDAEAGNIHYNGNQIGALLNLR 335
 ; 80 MAAMRGCGWLALDELHNVRLCFQ-----QSLHLEDEASFSDIVSGFIEHAAEVR 127
 ; 336 IKQTEG-----LSN-RIMFQISVSGGLAKSLAQYHNVNFKVLTGFKYIAAEIR 383
 ;
 ; RESULT 15
 ; US-09-543-681A-8138
 ; Sequence 8138, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:
 ; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 8138
 ; LENGTH: 572
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-8138
 ;
 ; Query Match 10.0%; Score 71.5; DB 4; Length 572;
 ; Best Local Similarity 20.9%; Pred. No. 3.6; Indels 58; Gaps 7;
 ; Matches 41; Conservative 32; Mismatches 58;
 ;
 ; 4 SQORVE----RFLQYFSGCKTPIHLKDGVCALYNEQ-----DEEAAVLEVPQ 48
 ; 125 SQEIRNLKGRFIEYFSAALNQLLEILLQDKNLFQEFQFTTSYQDGFYKYLAYYMRNEKF 184
 ; 49 SDSLLHLCRIIEADPQTSITILYSMLLQLNFEMAAEGCGWLALD----- 91

Db 185 NNQL---NSALEASHORSI--ITWVILAIIVVTALIIICWFALRNALIRPLNSLLNIKAF 239
QY 92 -----ELH--NVRLCFQOSLEHLDEA-----SFSDIVSGFIEHA 123
Db 240 SEGDLRPDIHVHGRNENMSLLASGLKHMQQELIHTVIRGVYQSTENIYNSTSEIAAGNNDLS 299
QY 124 AEVREYIAQLDESSAA 139
Db 300 SRTEEQVASLEETAAS 315

Search completed: July 7, 2004, 15:21:54
Job time : 3.67172 secs

Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:20:22 ; Search time 6.67931 Seconds
(without alignments)
6477.980 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	715	100.0	139	US-09-879-248-10	Sequence 10, Appl
2	715	100.0	139	US-10-441-736-10	Sequence 10, Appl
3	86.5	12.1	142	US-10-365-742-24	Sequence 24, Appl
4	82	11.5	792	US-10-282-122A-48692	Sequence 48692, A
5	81.5	11.4	402	US-10-282-122A-68346	Sequence 68346, A
6	80.5	11.3	766	US-10-664-456-12	Sequence 12, Appl
7	80.5	11.3	787	US-10-664-456-13	Sequence 13, Appl
8	80.5	11.3	820	US-10-664-456-14	Sequence 14, Appl
9	80.5	11.3	820	US-10-408-765A-1252	Sequence 1252, Ap
10	80	11.2	327	US-10-425-114-46300	Sequence 46300, A
11	79	11.0	327	US-10-424-599-239310	Sequence 239310,
12	79	11.0	789	US-09-712-363-244	Sequence 244, App
13	78.5	11.0	418	US-10-369-493-14001	Sequence 14001, A
14	78.5	11.0	830	US-10-369-493-17570	Sequence 17570, A
15	77	10.8	130	US-10-365-742-12	Sequence 12, Appl

16	76.5	10.7	1105	11	US-09-895-606-2	Sequence 2, Appl
17	75.5	10.6	966	9	US-09-801-368-372	Sequence 372, App
18	75	10.5	317	14	US-10-166-225A-85	Sequence 85, Appl
19	74	10.3	422	12	US-10-282-122A-69060	Sequence 69060, A
20	74	10.3	443	15	US-10-369-493-15331	Sequence 15331, A
21	74	10.3	443	15	US-10-369-493-15698	Sequence 15698, A
22	74	10.3	443	15	US-10-369-493-16089	Sequence 16089, A
23	73.5	10.3	376	16	US-10-437-963-183670	Sequence 183670,
24	73.5	10.3	984	14	US-10-354-358-102	Sequence 102, App
25	73.5	10.3	984	15	US-10-116-275-147	Sequence 147, App
26	73.5	10.3	1237	15	US-10-334-143-33	Sequence 33, Appl
27	73.5	10.3	2032	10	US-09-978-244A-12	Sequence 12, Appl
28	73	10.2	343	12	US-10-425-114-49498	Sequence 49498, A
29	72.5	10.1	421	12	US-10-282-122A-69899	Sequence 69899, A
30	72	10.1	961	16	US-10-437-963-163590	Sequence 163590,
31	71.5	10.0	362	12	US-10-424-599-174145	Sequence 174145,
32	71.5	10.0	422	16	US-10-437-963-120281	Sequence 120281,
33	71.5	10.0	511	15	US-10-104-047-2339	Sequence 2339, Ap
34	71.5	10.0	619	15	US-10-369-493-9885	Sequence 9885, Ap
35	71.5	10.0	920	10	US-09-919-039-76	Sequence 76, Appl
36	71.5	10.0	967	14	US-10-197-666A-90	Sequence 90, Appl
37	71.5	10.0	1170	16	US-10-437-963-120284	Sequence 120284,
38	71.5	10.0	1591	14	US-10-197-666A-92	Sequence 92, Appl
39	71	9.9	178	16	US-10-437-963-117593	Sequence 117593,
40	71	9.9	226	16	US-10-437-963-177193	Sequence 177193,
41	70.5	9.9	583	16	US-10-437-963-186085	Sequence 186085,
42	70.5	9.9	987	9	US-09-736-969A-15	Sequence 15, Appl
43	70.5	9.9	987	9	US-09-736-960-15	Sequence 15, Appl
44	70.5	9.9	987	9	US-09-736-968A-15	Sequence 15, Appl
45	70.5	9.9	987	12	US-09-737-246-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-879-248-10
; Sequence 10, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; FILE REFERENCE: 21829/A1
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-09-879-248-10

Query Match 100.0%; Score 715; DB 9; Length 139;
Best Local Similarity 100.0%; Pred. No. 5.2e-73;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTSSQQRVERFLQFSAGCKTPTHLKDGVCALYNEQDEEAALVLEVPQHSLLHLCKRIIE 60
Db 1 MTSSQQRVERFLQFSAGCKTPTHLKDGVCALYNEQDEEAALVLEVPQHSLLHLCKRIIE 60
QY 61 ADPQTSITLYSMLLQNFEMAMRGCMWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
Db 61 ADPQTSITLYSMLLQNFEMAMRGCMWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
QY 121 EHAAEVREYIAQLDESSAA 139
Db 121 EHAAEVREYIAQLDESSAA 139

25 LKDGVCALYNEQDEBAVLEVPQHSLSLLHCRITHEADPQTSITLYSMILQINFEMAMR 84
26 LRDDGYLLWQGDKQASLLVPSTGDALFAICTLSRVDPDCHGRLLALHLNLSVPHVM 85
85 GCWLALDELHNVRLCFQOSLEHLDEASPSDIVSGFTIEHAAVREYIAQLDES 136
86 SACTIALDVEQNT-LCURYT--HDLGGNGADTLLALENAQALAEQIKQVLEN 134

RESULT 4
US-10-282-122A-48692
; Sequence 48692, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48692
; LENGTH: 792
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48692

Query Match 11.5%; Score 82; DB 12; Length 792;
Best Local Similarity 22.5%; Pred. No. 3.8;
Matches 42; Conservative 27; Mismatches 42; Indels 76; Gaps 10;

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYN-----FLQYFSAGCKTPIHLKDGVCALYN----- 34
Db 535 MQNIQESVEREHLIVAKQMAEKALQSFNNNSHEIRTEL---NAIVGFTNVLLGEGS 591
QY 35 ---EQDEAAVLEVPQHSLSLLHCRITHEADPQTSITLYSMILQ-----LNFEMAMR 83
Db 592 BEIDPDEKASMLEITINHNNELL-----KLINDVLEIRSLDSGLDFDMKE- 637
QY 84 RGCWLALD---ELHN-----VRLCFQOSLE-----HLDEASPSDIVSGFTIEHAAV 126

25 LKDGVCALYNEQDEBAVLEVPQHSLSLLHCRITHEADPQTSITLYSMILQINFEMAMR 84
26 LRDDGYLLWQGDKQASLLVPSTGDALFAICTLSRVDPDCHGRLLALHLNLSVPHVM 85
85 GCWLALDELHNVRLCFQOSLEHLDEASPSDIVSGFTIEHAAVREYIAQLDES 136
86 SACTIALDVEQNT-LCURYT--HDLGGNGADTLLALENAQALAEQIKQVLEN 134

RESULT 2
US-10-441-736-10
; Sequence 10, Application US/10441736
; Publication No. US20040016029A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/203 (BBC-003)
; CURRENT APPLICATION NUMBER: US/10/441,736
; CURRENT FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/107,243
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 09/431,614
; PRIOR FILING DATE: 1999-11-02
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 139
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-10-441-736-10

Query Match 100.0%; Score 715; DB 15; Length 139;
Best Local Similarity 100.0%; Pred. No. 5.2e-73; Indels 0; Gaps 0;
Matches 139; Conservative 0; Mismatches 0

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEBAVLEVPQHSLSLLHCRIT 60
Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEBAVLEVPQHSLSLLHCRIT 60
QY 61 ADPQTSITLYSMILQINFEMAMRGWCWLALDELHNVRLCFQOSLEHLDEASPSDIVSGFI 120
Db 61 ADPQTSITLYSMILQINFEMAMRGWCWLALDELHNVRLCFQOSLEHLDEASPSDIVSGFI 120
QY 121 EHAAVREYIAQLDESSAA 139
Db 121 EHAAVREYIAQLDESSAA 139

RESULT 3
US-10-365-742-24
; Sequence 24, Application US/10365742
; Publication No. US20030204868A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Cartinhour, Samuel W.
; APPLICANT: Schneider, David J.
; APPLICANT: Tang, Xiaoyan
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING
; TITLE OF INVENTION: NUCLEIC ACIDS, AND USE THEREOF
; FILE REFERENCE: 19603/4112
; CURRENT APPLICATION NUMBER: US/10/365,742
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/356,408
; PRIOR FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: 60/380,185
; PRIOR FILING DATE: 2002-05-10
; NUMBER OF SEQ ID NOS: 209
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato DC3000
US-10-365-742-24

Query Match 12.1%; Score 86.5; DB 12; Length 142;
Best Local Similarity 25.9%; Pred. No. 0.12;
Matches 29; Conservative 23; Mismatches 57; Indels 3; Gaps 2;

Db 638 ---WNMTDIVKEIKYQPLRLSLQRLLEDDTVSPVHTDLRFVQVSNFLNNKPF 694
 QY 127 RE--YIA 131
 Db 695 TQGYIA 701

RESULT 5
 US-10-282-122A-68346
 ; Sequence 68346, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 68346
 ; LENGTH: 402
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas putida
 US-10-282-122A-68346

Query Match 11.4%; Score 81.5; DB 12; Length 402;
 Best Local Similarity 20.5%; Pred. No. 1.8;
 Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;
 QY 12 LQVFSACKTPI--HLKDGVCALYNEQDEAAVLEVPQHSD-----VSQADLDKAWNAFNAKTYRGIC 239
 Db 192 IERISRDARVPVHKHLDGICHY-----VSQADLDKAWNAFNAKTYRGIC 239
 QY 51 ----SLLLHCRILIEADQTSITLTYMLQLNLFENAAVRGC-----WLA- 89
 Db 240 GAMETLVDQVABR-----FLPEMARRFVEKGVLELGCRTQAIISAKPATEADWHT 293
 QY 90 -LDELHNVRLC--PQGSLEHLEDEASFSFDSFTEHAAAEVREYIAQLDESS 137
 Db 294 YLDAILSIKRVVDGLNQLAIEHINHYGSHHTDSIISEHOGEARQFMAEVDAS 344

RESULT 6
 US-10-664-456-12
 ; Sequence 12, Application US/10664456
 ; Publication No. US20040038364A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Immunex Corporation
 ; TITLE OF INVENTION: NEW METALLOPROTEINASE-DISINTEGRIN FAMILY MEMBERS: SVPH
 ; FILE REFERENCE: 03260.0093-00304
 ; CURRENT APPLICATION NUMBER: US/10/664,456
 ; CURRENT FILING DATE: 2003-09-19
 ; PRIOR APPLICATION NUMBER: prior APPLICATION NUMBER: US/09/890,323
 ; PRIOR FILING DATE: prior FILING DATE: 2001-07-25
 ; PRIOR APPLICATION NUMBER: 60/116,670
 ; PRIOR FILING DATE: 1999-01-21
 ; PRIOR APPLICATION NUMBER: 60/138,682
 ; PRIOR FILING DATE: 1999-06-14
 ; PRIOR APPLICATION NUMBER: 60/155,798
 ; PRIOR FILING DATE: 1999-09-27
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 12
 ; LENGTH: 766
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-664-456-12

Query Match 11.3%; Score 80.5; DB 12; Length 766;
 Best Local Similarity 22.3%; Pred. No. 5.4;
 Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;
 QY 18 GCKTPIHLK-----DGVCALYNEQDEAAVLEVPQHSDSLLLHCRILIEADQTSITL-- 69
 Db 60 GQKHLIHKVKKLLFSKHLPVFTYDQGAILEDQFVQNNCYH-GYVEGDPESLVSLS 118
 QY 70 ----YSMLLQLNFEMAAVRGCWLALDELHNVRLCFOOSLEHL-----DEASFSFDSVSG 118
 Db 119 CFGGFQGIQLN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165
 QY 119 FIEHAAEVREYIAQLDESS 137
 Db 166 FMQNEITCRMEFEIDNST 184

RESULT 7
 US-10-664-456-13
 ; Sequence 13, Application US/10664456
 ; Publication No. US20040038364A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Immunex Corporation
 ; TITLE OF INVENTION: NEW METALLOPROTEINASE-DISINTEGRIN FAMILY MEMBERS: SVPH
 ; FILE REFERENCE: 03260.0093-00304
 ; CURRENT APPLICATION NUMBER: US/10/664,456
 ; CURRENT FILING DATE: 2003-09-19
 ; PRIOR APPLICATION NUMBER: prior APPLICATION NUMBER: US/09/890,323
 ; PRIOR FILING DATE: prior FILING DATE: 2001-07-25
 ; PRIOR APPLICATION NUMBER: 60/116,670
 ; PRIOR FILING DATE: 1999-01-21
 ; PRIOR APPLICATION NUMBER: 60/138,682
 ; PRIOR FILING DATE: 1999-06-14
 ; PRIOR APPLICATION NUMBER: 60/155,798
 ; PRIOR FILING DATE: 1999-09-27
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 13
 ; LENGTH: 787
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-664-456-13

Query Match 11.3%; Score 80.5; DB 12; Length 787;

Best Local Similarity 22.3%; Pred. No. 5.6; Mismatches 28; Conservative 31; Gaps 6; Indels 47; Mismatches 47; Indels 33; Gaps 6;

18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIIEADPQTSITL-- 69
60 GQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLSLST 118
70 -----YSMLLQINFEMAAMRGCLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
119 CFQGFQGILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165
119 FIEHAAEVREYIAQLDESS 137
166 FMQNEITCRMEFEIDNST 184

RESULT 8
US-10-664-456-14
Sequence 14, Application US/10664456
Publication No. US20040038364A1
GENERAL INFORMATION:
APPLICANT: Immunex Corporation
TITLE OF INVENTION: NEW METALLOPROTEINASE-DISINTEGRIN FAMILY MEMBERS: SVPH
TITLE OF INVENTION: DNAS AND POLYPEPTIDES
FILE REFERENCE: 03260.0093-00304
CURRENT APPLICATION NUMBER: US/10/664,456
CURRENT FILING DATE: 2003-09-19
PRIOR APPLICATION NUMBER: prior application number: US/09/890,323
PRIOR FILING DATE: prior filing date: 2001-07-25
PRIOR APPLICATION NUMBER: 60/116,670
PRIOR FILING DATE: 1999-01-21
PRIOR APPLICATION NUMBER: 60/138,682
PRIOR FILING DATE: 1999-06-14
PRIOR APPLICATION NUMBER: 60/155,798
PRIOR FILING DATE: 1999-09-27
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 14
LENGTH: 820
TYPE: PRT
ORGANISM: Homo sapiens
US-10-664-456-14

Query Match 11.3%; Score 80.5; DB 12; Length 820;
Best Local Similarity 22.3%; Pred. No. 5.9; Mismatches 28; Conservative 31; Gaps 6; Indels 47; Mismatches 47; Indels 33; Gaps 6;

18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIIEADPQTSITL-- 69
60 GQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLSLST 118
70 -----YSMLLQINFEMAAMRGCLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
119 CFQGFQGILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165
119 FIEHAAEVREYIAQLDESS 137
166 FMQNEITCRMEFEIDNST 184

TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1252
LENGTH: 820
TYPE: PRT
ORGANISM: Homo sapiens
US-10-408-765A-1252

Query Match 11.3%; Score 80.5; DB 16; Length 820;
Best Local Similarity 22.3%; Pred. No. 5.9; Mismatches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIIEADPQTSITL-- 69
60 GQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLSLST 118
70 -----YSMLLQINFEMAAMRGCLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
119 CFQGFQGILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165
119 FIEHAAEVREYIAQLDESS 137
166 FMQNEITCRMEFEIDNST 184

RESULT 10
US-10-425-114-46300
Sequence 46300, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 46300
LENGTH: 327
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 700764472_FLI.pep
US-10-425-114-46300

Query Match 11.2%; Score 80; DB 12; Length 327;
Best Local Similarity 30.8%; Pred. No. 2; Mismatches 21; Conservative 44; Indels 48; Mismatches 30; Gaps 10;

24 HKDGVCALYNEQDEE-----AAV--LEVPQHSLSLLHCRRIIEADPQT-----SIT 68
16 YLENGSVEIYSRHAERTGKFPDVVAASRLKPKTVS-SLILDCEIVAYDQTQTIHSFQ 74
69 LYSMLLQINFEMAAMR--GCWLALDELH-NVRLCFQQSL-----BHLDEASFSDIVSGFIE 121
75 ALSTRAKNVEMEDIKVDVCIFAFDLYINGQALIQENLRVRREHL-VASFEE-EPGFLO 132
122 HAA-----EVREYIAQLDESSAA 139
133 FATTITSDNVEIEIKFLDQAVCA 155

RESULT 11
US-10-424-599-239310

TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:14:07 ; Search time 2.3049 Seconds
(without alignments)

5593.254 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	100.0	139	2 T18449	pathogenicity fact
2	273.5	38.3	129	2 T30333	avirulence protein
3	79	11.0	789	2 A70868	probable plb2 - M
4	78.5	11.0	830	2 C82710	ribonucleoside-dip
5	77.5	10.8	775	2 T45238	probable transfera
6	75.5	10.6	402	2 T19390	hypothetical prote
7	75.5	10.6	966	2 S25365	CyC8 protein - yea
8	75	10.5	317	2 D95044	diphosphomevalonat
9	73.5	10.3	984	1 A34076	protein-tyrosine k
10	73	10.2	1708	2 AB1866	WD-40 repeat prote
11	72.5	10.1	131	2 AG2034	hypothetical prote
12	71.5	10.0	1036	2 T31673	N-acetylglucosamin
13	71.5	10.0	2484	2 T26216	hypothetical prote
14	71.5	10.0	2607	2 T26215	hypothetical prote
15	71	9.9	637	2 B95878	probable adenylate
16	70.5	9.9	1048	2 T31653	hypothetical prote
17	69.5	9.7	291	2 T23051	hypothetical prote
18	69.5	9.7	393	2 S76366	hypothetical prote
19	69.5	9.7	2183	1 ZLN2MV	genome polyprotein
20	69.5	9.7	2183	1 G48556	genome polyprotein
21	69	9.7	225	2 F72124	pts iia protein +
22	69	9.7	344	2 C97914	diphosphomevalonat
23	69	9.7	374	2 A42264	membrane-associate
24	69	9.7	451	1 F64155	hypothetical prote
25	69	9.7	610	2 T02298	probable outer den
26	69	9.7	638	2 T09400	outer dense fiber
27	69	9.7	727	2 C84748	hypothetical prote
28	68.5	9.6	130	2 AG2226	hypothetical prote
29	68.5	9.6	165	2 C72625	hypothetical prote

30 68.5 9.6 167 2 H81717 conserved hypothet
31 68.5 9.6 173 2 S75578 hypothetical prote
32 68.5 9.6 352 2 T44968 gas-vesicle operon
33 68.5 9.6 591 2 I54368 merlin protein - m
34 68.5 9.6 595 2 S33809 neurofibromin 2 -
35 68.5 9.6 596 2 T68664 merlin - mouse
36 68.5 9.6 621 2 T06717 hypothetical prote
37 68.5 9.6 794 2 T46073 hypothetical prote
38 68.5 9.6 883 2 A96805 hypothetical prote
39 68 9.5 178 2 T39722 ORF12 - Agrobacter
40 68 9.5 225 2 F86498 Pts IIA protein wi
41 68 9.5 478 2 G81679 3-dehydroquinat d
42 68 9.5 638 2 T03791 outer dense fiber
43 68 9.5 923 2 T24712 hypothetical prote
44 68 9.5 1119 2 B70126 surface-located me
45 68 9.5 1144 2 A54810 TMV resistance pro

ALIGNMENTS

RESULT 1

T18449

Pathogenicity factor DspB - Erwinia amylovora

C;Species: Erwinia amylovora

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C;Accession: T18449

R;Gaudriault, S.

submitted to the EMBL Data Library, May 1998

A;Reference number: Z18936

A;Accession: T18449

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-139 <GAU>

A;Cross-references: EMBL:Y13831; PIDN:CAA74157.1

A;Experimental source: strain CfBP1430; specific host Pommoideae

C;Genetics:

A;Note: dspB

C;Function:

A;Description: involved in pathogenicity

Query Match 100.0%; Score 715; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.8e-64;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAALVLEVPQHSLLHCRITIE 60

Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAALVLEVPQHSLLHCRITIE 60

Qy 61 ADPQTSITLYSMLLQLNFEMAAAGCWLALDELHNRLCFOQSLHLDDEAFSDIVSGFI 120

Db 61 ADPQTSITLYSMLLQLNFEMAAAGCWLALDELHNRLCFOQSLHLDDEAFSDIVSGFI 120

Qy 121 EHAAEVREYIAQLDESSAA 139

Db 121 EHAAEVREYIAQLDESSAA 139

RESULT 2

T30333

avirulence protein - Pseudomonas syringae

C;Species: Pseudomonas syringae

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C;Accession: T30333

R;Bogdanov, A.J.; Kim, J.P.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.; Co

Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998

A;Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dsrEF,

A;Reference number: Z20825; MUID:98115919; PMID:9448330

A;Accession: T30333

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-129 <BOG>

A;Cross-references: EMBL:U97505; NID:g2978502; PID:g2978504; PIDN:AAC06135.1

C:Genetics:
A:Gene: avrF

Query Match 38.3%; Score 273.5; DB 2; Length 129;
Best Local Similarity 43.1%; Pred. No. 4.1e-20;
Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 3;

QY 1 MTSSQORVERFLOYFAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLLHLCRIIE 60
Db 1 MKTSQDPFAFFINSLGAQLGTSITLQNGVCALYDQGNNEAAITELPHESEMIFHCRIGR 60

QY 61 ADPQTSITLYSMLLQNFEMAMRGWLALDELHNRLCFOQSLHLEDEASFSDIYSGFI 120
Db 61 C-PERAPDLL-LRLSLNFDVARLHGCVFAVDQ-GDVRCLCAQREBLASLDEPAFCVTRGFI 117

QY 121 EHAAEVREXI 130
Db 118 SQAREARAF 127

RESULT 3
A70868
probable plsB2 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 12-Jun-2003
C:Accession: A70868
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
C:Accession: A70868
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-789 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16059.1; PID:e123760
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: plsB2
C:Superfamily: glycerol-3-phosphate O-acyltransferase

Query Match 11.0%; Score 79; DB 2; Length 789;
Best Local Similarity 24.3%; Pred. No. 8.7;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

QY 23 IHLKDGVCALYNEQDEAAVLEVPQHSLSLLHLCRIIEADPQTSITLYSMLLQNFEMAA 82
Db 445 VSMRQYLGAHGELTQDPAARLQALQKNSFEVAMRILQATPVTATGLVSALL-----LT 498

QY 83 MRGWLALDELHNRLCFOQSLHLD-----EASFSDIYSGF--- 119
Db 499 TRGTALTLDQLHHT---LQSLDYLERKQSPVSTSLRLRSRCVRAADALSNHGHPVTR 555

QY 120 IEHAAEVREXIQAQDESSAA 139
Db 556 VDSGREGPVWYIAPDDEHAAA 575

RESULT 4
C82710
ribonucleoside-diphosphate reductase alpha chain XF1196 [imported] - Xylella fastidiosa
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: C82710
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: C82710
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-830 <SIM>
A:Cross-references: GB:AE003954; GB:AE003849; NID:g9106165; PIDN:AAF84006.1; GSPDB:GN0001
A:Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, P.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canargo, L.E.A.; Carraro, D.M.; Carrer, F. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohman, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, M.A.; Martins, E.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.M.B.N.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.C.R.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaka, A.; da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z. A:Reference number: A59328
A:Contents: annotation
A:Gene: XF1196
C:Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain

Query Match 11.0%; Score 78.5; DB 2; Length 830;
Best Local Similarity 25.6%; Pred. No. 10;
Matches 33; Conservative 21; Mismatches 46; Indels 29; Gaps 6;

QY 17 AGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHLCRII-EADPQTSITLYSMMLQ 75
Db 22 AGSVPPVALS---VFHPDEPD---EVPIPQTMHTRAVEANVSTWITKEAGNRR 73

QY 76 LNFEMAMRGWLALDELHNRLCFOQSLHLEDEASFSDIYSGFIEHAAEV----- 126
Db 74 MPDFCARLE---CAIDTII-----QEPQLDVAEYKRAVGFVERKSVNADDLVLL 123

QY 127 -REVIAQLD 134
Db 124 IREAEARVD 132

RESULT 5
T45238
probable transferase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 12-Jun-2003
C:Accession: T45238
R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, March 1999
A:Reference number: Z22949
A:Accession: T45238
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-775 <JAM>
A:Cross-references: EMBL:AL049913; PIDN:CAB43153.1
A:Experimental source: cosmid B1610
C:Genetics:
A:Note: plsB
C:Superfamily: glycerol-3-phosphate O-acyltransferase

Query Match 10.8%; Score 77.5; DB 2; Length 775;
Best Local Similarity 25.8%; Pred. No. 12;
Matches 34; Conservative 16; Mismatches 33; Indels 49; Gaps 5;

QY 11 FLOYFSAGCKTPIHLKDGVCALYN-----EQ 36
Db 397 YAAVARGGKTP---EGVAWLYSFTKAQGRNVGKIYVRFPEAVSMRQYLGAPHGALVQ 452

QY 37 DEAAVLEVPQHSLSLLHLCRIIEADPQTSITLYSMMLQNFEMAMRGWLALDELHN 96
Db 453 DQDAKRLALQKMSFEVA--WRILCATPVTATVLSALL-----LITRGVALTLDQLHHT 504

QY 97 RLCFQOQSLHLD 108
Db 505 ---LQESLDYLE 513

RESULT 6

TI9390

hypothetical protein Cl8D4.2 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19390

R:Mortimore, B.

Submitted to the EMBL Data Library, November 1996

A:Reference number: Z19119

A:Accession: T19390

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-402 <WIL>

A:Cross-references: EMBL:Z81474; PIDN:CA803900.1; GSPDB:GN00023; CESP:Cl8D4.2

A:Experimental source: clone Cl8D4

C:Genetics:

A:Gene: CESP:Cl8D4.2

A:Map position: 5

A:Introns: 326/3

Query Match

Best Local Similarity 10.6%; Score 75.5; DB 2; Length 402;

Matches 33; Conservative 23; Mismatches 35; Indels 45; Gaps 8;

Qy

5 QORVERFLOYF-----SAGCKTPHLKDGVCALYNEQDEAAVLEVPQHSLSLL--HC 56

Db

202 EQRIDLFVKYFEDILSSVGHVTPVE-----ALYMQACNQTVLVLRLHLSRLKTIISI 255

Qy

57 RIIEADPQTSITLYSMLLQNFEMFAAMRGCMALDELHNRLCFQOSLEHLDEASFSDIV 116

Db

256 KIVE-DDQDKPCNVLPLEIN-----BIC-----ELDQWKFAENI 289

Qy

117 SGFTIEH---AAEVREY 129

Db

290 E--VEHLIDTDVVRHF 303

RESULT 7

S25365

CYC8 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: glucose repression mediator; protein YBR112c; SSN6 p

C:Species: *Saccharomyces cerevisiae*

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000

C:Accession: S25365; S48277; S45980; S25404; S25405; A30306; S44692

R:Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.

Yeast 8, 397-408, 1992

A:Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excision i

A:Reference number: S25364; MUID:92327848; PMID:1626431

A:Accession: S25365

A:Molecule type: DNA

A:Residues: 1-966 <MAN>

A:Cross-references: EMBL:X66247; NID:g3548; PIDN:CAA46973.1; PID:g3550

R:Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:7900426

A:Accession: S48277

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-966 <MAN>

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55615.1; PID:g476068

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994

R:Feldmann, H.; Mannhaupt, G.; Schwarzlöse, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45927

A:Accession: S45980

A:Molecule type: DNA

A:Residues: 1-966 <FE2>

A:Cross-references: EMBL:Z35981; NID:g536449; PIDN:CAA85069.1; PID:g536450; MIPS:YBR112c

R:Schultz, J.; Carlson, M.

Mol. Cell. Biol. 7, 3637-3645, 1987

A:Title: Molecular analysis of SSN6, a gene functionally related to the SNF1 protein kinase

A:Reference number: S25404; MUID:88065502; PMID:3316983

A:Accession: S25404

A:Molecule type: DNA

A:Residues: 1-546, 'K', 548-966 <SCH>

A:Cross-references: EMBL:M17826; NID:gl72725; PIDN:AAA35103.1; PID:gl72726

R:Trumbly, R.J.

Gene 73, 97-111, 1988

A:Title: Cloning and characterization of the CYC8 gene mediating glucose repression in *Y*

A:Reference number: S25405; MUID:89211964; PMID:2854095

A:Accession: S25405

A:Molecule type: DNA

A:Residues: 1-546, 'K', 548-966 <TRU>

A:Cross-references: EMBL:M23440; NID:gl71349; PIDN:AAA34545.1; PID:gl71350

C:Genetics:

A:Gene: SGD:CYC8; SSN6; CRT8

A:Cross-references: SGD:S0000316; MIPS:YBR112c

A:Map position: 2R

C:Function:

A:Description: required for complete derepression of ICL1; required for repression of SUC

C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat hom

C:Keywords: nucleus; transcription regulation

F:24-257/Domain: tetratricopeptide repeat homology <TT1>

F:262-295/Domain: tetratricopeptide repeat homology <TT2>

F:296-329/Domain: tetratricopeptide repeat homology <TT3>

F:330-363/Domain: tetratricopeptide repeat homology <TT4>

F:365-398/Domain: tetratricopeptide repeat homology <TT5>

Query Match 10.6%; Score 75.5; DB 2; Length 966;

Best Local Similarity 26.2%; Pred. No. 25;

Matches 22; Conservative 15; Mismatches 34; Indels 13; Gaps 2;

Qy

64 QTSITLYSMLLQNFEM-----AAMRGCMALDELHNRLCFQOSLEHLDEASFSDIVSGF 119

Db

97 QRAAELVERALLVNPPLSDVWATLGHCVLMDDQLQRAYNQALYHLNPNVPLMWHGI 156

Qy

120 -----IEHAAEVREYIAQLD 134

Db

157 GILYDRYGLDYAEFAFAKVLIED 180

RESULT 8

D95044

diphosphomevalonate decarboxylase [imported] - *Streptococcus pneumoniae* (strain TIGR4)

C:Species: *Streptococcus pneumoniae*

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 18-Aug-2003

C:Accession: D95044

R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: D95044

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-317 <KUR>

A:Cross-references: GB:AB005672; PIDN:AAK74549.1; PID:gl4971852; GSPDB:GN00164; TIGR:SP4:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0382

C:Superfamily: diphosphomevalonate decarboxylase

Query Match 10.5%; Score 75; DB 2; Length 317;

Best Local Similarity 23.8%; Pred. No. 7.2;

Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;

Qy

20 KTIHLKDG--VCA-----LYNEQDEAAVLEVPQHSLSLLHC 56

Db

180 KPISSRGMKLCVETSTTFDDVVRQSEKDYQDMILYLKENDFAKIGELTE-KNALAMH- 237

Qy

57 RIIEADPQTSITLYSMLLQNFEMFAAMRGCMALDELHNRLCFQOSLEHL 107

Db 238 ----ATTKTAPAFSYLTASAEAMAFVQLRKEAGECYFTMDAGPNVKVFCQEKDLEHL 293
 QY 108 DE 109
 Db 294 SE 295

RESULT 9

A34076
 N;Alternate names: protein-tyrosine kinase (EC 2.7.1.112) receptor type eph 1 precursor - human
 C;Species: Homo sapiens (man)
 C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Feb-2000
 C;Accession: A34076; S44280
 R;Hirai, H.; Maru, Y.; Hagiwara, K.; Nishida, J.; Takaku, F.
 Science 238, 1717-1720, 1987
 A;Title: A novel putative tyrosine kinase receptor encoded by the eph gene.
 A;Reference number: A34076; MUID:88070650; PMID:2825356
 A;Accession: A34076
 A;Molecule type: mRNA
 A;Residues: 1-984 <HIR>
 A;Cross-references: GB:M18391; NID:G3397116; PIDN:AAA36747.1; PID:G339717
 A;Note: the sequence in GenBank entry HUMTKR, release 111.0, has the codons GCG for 398-
 R;Tuzi, N.L.
 submitted to the EMBL Data Library, November 1993

A;Description: An EGFR/eph chimeric receptor possesses ligand stimulated tyrosine kinase
 A;Reference number: S44280
 A;Accession: S44280
 A;Molecule type: mRNA
 A;Residues: 286-397, A;399-580, 'ORDRATVDREKWLKPYVDLQAYEDPAQALDF', 583,625-984 <TUZ>
 A;Cross-references: EMBL:Z27409; NID:G482916; PIDN:CAA81796.1; PID:G482917
 C;Genetics:
 A;Gene: GDB:EPHT1; EPH; EPHT
 A;Cross-references: GDB:119875; OMIM:179610
 A;Map position: 7q32-7q36
 C;Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat h
 F;1-23/Domain: signal sequence #status predicted <SIG>
 F;24-984/Product: protein-tyrosine kinase receptor type eph 1 #status predicted <MAY>
 F;548-568/Domain: transmembrane #status predicted <TM>
 F;630-895/Domain: protein kinase homology <KIN>
 F;638-646/Region: protein kinase ATP-binding motif
 F;918-984/Domain: SAM homology <SAM>
 F;59,338,414,478/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 10.3%; Score 73.5; DB 1; Length 984;
 Best Local Similarity 27.5%; Pred. No. 40;
 Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6;

QY 7 RVERFLQYFAGCKT-----PIHLKDGVCALYNEQDEAAVLEVPQHSDSLLLHCRHIE 60
 Db 94 RVHVELQFTVRDCCKSPFGAGPLGCKETFNLLYMSDQDVGI---QLRRLFLQKVTVA 149
 QY 61 ADPQTSI-TLYSMLLQINFEMAM-----RCGWLADDE-----LHNVRLCFQOSLEHLD 108
 Db 150 ADQSFTRDLASGVKLVNRCVSLGRLTRGLYLAHPGACVALYSVRVYQRCPTLN 209
 QY 109 E-ASFSFDIVSG 118
 Db 210 GLAQFFDTLPG 220

RESULT 10

AE1866
 WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
 C;Species: Nostoc sp. PCC 7120
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C;Accession: AE1866
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AE1866
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1708 <KUR>
 A;Cross-references: GB:BA000019; PIDN:BA072436.1; PID:gl7129823; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: all0478

Query Match 10.2%; Score 73; DB 2; Length 1708;
 Best Local Similarity 30.2%; Pred. No. 87;
 Matches 26; Conservative 11; Mismatches 35; Indels 14; Gaps 2;

QY 59 IEADPQTSITLYSMLLQINFEMAMRCGWLADDELHNV-----RLCFQOSLEHLD 109
 Db 276 VRPTFLSSPIRLQGVLLQYIGL-----CYCLAEQNQLDNRHWHETAKFYQECLEILQV 330
 QY 110 ASFSFDIVSGFIEHAAEVREYIAQLDE 135
 Db 331 AGRPDIASEFIGQLAEVLEHLQAWDE 356

RESULT 11

AG2034
 hypothetical protein all1829 [imported] - Nostoc sp. (strain PCC 7120)
 C;Species: Nostoc sp. PCC 7120
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C;Accession: AG2034
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AG2034
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-131 <KUR>
 A;Cross-references: GB:BA000019; PIDN:BA073528.1; PID:gl7130919; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: all1829

Query Match 10.1%; Score 72.5; DB 2; Length 131;
 Best Local Similarity 23.6%; Pred. No. 4.4;
 Matches 25; Conservative 20; Mismatches 46; Indels 15; Gaps 3;

QY 24 HLKDG---VCALYNEQDEAAVLEVPQHSDSLLLHCRHIEADPQTSITLYSMLLQINFEM 80
 Db 31 HLEQSDYKVCYGVWDEODKYVEKITLPRPLETELVSISI-----GVTHTERFLQKFSL 83
 QY 81 AAMRGCGWLADDELHNV-----LCFQOSLEHLDASFSDIVSGFIE 121
 Db 84 RARFTYAIQAQNNHTQKIGELVLVYDENLEFIDENWLLDVSMPLE 129

RESULT 12

T31673
 N-acetylglucosaminyltransferases (EC 2.4.1.-), chain p110 - rat
 N;Alternate names: O-GlcNAc transferase
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999

C;Accession: T31673
 R;Kreppel, L.K.; Blomberg, M.A.; Hart, G.W.
 J. Biol. Chem. 272, 9308-9315, 1997
 A;Title: Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characteri
 A;Reference number: Z21056; MUID:97238869; PMID:9083067
 A;Accession: T31673
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1036 <KRE>

A;Cross-references: EMBL:U76557; NID:g1931578; PID:g1931579; PIDN:AAC53121.1
A;Experimental source: strain Sprague Dawley; liver
C;Genetics:
A;Gene: OGT
C;Keywords: glycosyltransferase; hexosyltransferase; tandem repeat

Query Match 10.0%; Score 71.5; DB 2; Length 1036;
Best Local Similarity 24.4%; Pred. No. 67;
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

Qy 24 HLKDGVCALYNEQDEEAALVLEVPQHSDSLHLCRIIEADPQT 70
Db 281 HFDPAYCNALKEGSAEAECDYNALRLCPHADSLNLANIKREQCNIEAVRL 340
Qy 71 SMLQLNFMAAMRGWLALDELHNRLCFQOSLEHLDEA-----SFSDIVSGFIEHAAE 125
Db 341 RKALEVPEFAAHS-----NLASVLQOQKLOEALMHYKEAIRISFTPADAYSNGNLTKE 397
Qy 126 VRE 128
Db 398 MQD 400

RESULT 13
T26216
hypothetical protein W06A7.3c - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26216
R;Ainscough, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z20173
A;Accession: T26216
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2484 <MIL>
A;Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GN00023; CESP:W06A7.3c
A;Experimental source: clone W06A7
C;Genetics:
A;Gene: CESP:W06A7.3c
A;Map position: 5
A;Introns: 1827/1; 1866/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match 10.0%; Score 71.5; DB 2; Length 2484;
Best Local Similarity 24.2%; Pred. No. 1.9e+02;
Matches 31; Conservative 19; Mismatches 43; Indels 35; Gaps 6;

Qy 18 GCKTPHUKDGCALYNEQDEEA-----AVLEVPQHSDSLHLCRIIEADPQT 65
Db 1491 GCITDVADSD-----VNEQDEESTLKILKVPSPSLELDFNDPKVHVPIPLMEPAT 1545
Qy 66 SITLYSMLQLNFMAAMRGWLALDELHNRLCFQOSLEHLDEASFSDIVSGFIEHAAE 125
Db 1546 -----MYLE-----WIIADAVKEV-----SEMEVVTESEISEMAPQVSESTCP 1587
Qy 126 VREYIAQL 133
Db 1588 IPEPLADL 1595

RESULT 14
T26215
hypothetical protein W06A7.3a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26215
R;Ainscough, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z20173
A;Accession: T26215
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2607 <MIL>

A;Cross-references: EMBL:Z78066; PIDN:CAB51522.2; GSPDB:GN00023; CESP:W06A7.3a
A;Experimental source: clone W06A7
C;Genetics:
A;Gene: CESP:W06A7.3a
A;Map position: 5
A;Introns: 1827/1; 1866/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2

Query Match 10.0%; Score 71.5; DB 2; Length 2607;
Best Local Similarity 24.2%; Pred. No. 2e+02;
Matches 31; Conservative 19; Mismatches 43; Indels 35; Gaps 6;

Qy 18 GCKTPHUKDGCALYNEQDEEA-----AVLEVPQHSDSLHLCRIIEADPQT 65
Db 1491 GCITDVADSD-----VNEQDEESTLKILKVPSPSLELDFNDPKVHVPIPLMEPAT 1545
Qy 66 SITLYSMLQLNFMAAMRGWLALDELHNRLCFQOSLEHLDEASFSDIVSGFIEHAAE 125
Db 1546 -----MYLE-----WIIADAVKEV-----SEMEVVTESEISEMAPQVSESTCP 1587
Qy 126 VREYIAQL 133
Db 1588 IPEPLADL 1595

RESULT 15
B95878
probable adenylate cyclase (EC 4.6.1.1) [imported] - Sinorhizobium meliloti (strain 1021)
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
C;Accession: B95878
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: B95878
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-637 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC48690.1; PID:g15140162; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: cyaf7; SM520300
A;Genome: plasmid
C;Keywords: phosphorus-oxygen lyase

Query Match 9.9%; Score 71; DB 2; Length 637;
Best Local Similarity 24.0%; Pred. No. 42;
Matches 29; Conservative 14; Mismatches 36; Indels 42; Gaps 5;

Qy 31 ALY-----NEQDE--EAAVLEVPQHSDSLH-----CRIIEADPQTS 66
Db 534 ALYAAGRLDEADVLRECLIRAPQADCLLIRTAVLVSQGVGAGQRTWARLTVDPDPE-- 591
Qy 67 ITLYSMLQLNFMAAMRGWLALDELHNRLCFQOSLEHLDEASFSDIVSGFIEHAAE 126
Db 592 -----FSLASERSMRFGDS-----ALMBQFLSRLAANAPDVTSGFLHPTQS 635
Qy 127 R 127
Db 636 R 636

Search completed: July 7, 2004, 15:21:06
Job time : 4.39049 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:07:11 ; Search time 1.96864 Seconds
(without alignments)
3676.525 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81.5	11.4	423	1	PROA_PSEPK
2	80.5	11.3	820	1	AD29 HUMAN
3	79	11.0	789	1	PLSB MYCTU
4	77.5	10.8	775	1	PLSB MYCLE
5	75.5	10.6	966	1	SSN6 YEAST
6	73.5	10.3	320	1	DP44 BR969
7	73.5	10.3	741	1	DOC8 MOUSE
8	73.5	10.3	976	1	EPAL HUMAN
9	72.5	10.1	421	1	PROA_PSEPM
10	71.5	10.0	1036	1	OGT1 HUMAN
11	71.5	10.0	1036	1	OGT1 RAT
12	71	9.9	816	1	NEL2 MOUSE
13	70.5	9.9	1799	1	DOC8 HUMAN
14	70.5	9.9	2203	1	POLG EC09B
15	70	9.8	946	1	AMPN PLUXY
16	70	9.8	1739	1	DOTL HUMAN
17	69.5	9.7	2183	1	RRPL MEASA
18	69.5	9.7	2183	1	RRPL MEASE
19	69.5	9.7	3680	1	DMD CANFA
20	69	9.7	451	1	YUN HAEIN
21	68.5	9.6	173	1	YCF3 SYN3
22	68.5	9.6	586	1	MERL RAT
23	68.5	9.6	595	1	MERL HUMAN
24	68.5	9.6	595	1	MERL PAPAN
25	68.5	9.6	596	1	MERL MOUSE
26	68	9.5	478	1	ARDE CHLMU
27	68	9.5	727	1	DOC7 MOUSE
28	68	9.5	1302	1	DOC7 HUMAN
29	67.5	9.4	404	1	O67C DROME
30	67.5	9.4	1141	1	GSX3 CAEEL
31	67.5	9.4	1527	1	MRP3 HUMAN
32	67	9.4	172	1	MUPL HUMAN
33	67	9.4	532	1	UD14_RABIT

34	67	9.4	578	1	YQD6 CAEEL
35	67	9.4	763	1	CADQ MOUSE
36	66.5	9.3	241	1	PCYA PROMP
37	66.5	9.3	283	1	DMPD_PSESP
38	66.5	9.3	412	1	NEMO MOUSE
39	66.5	9.3	756	1	EFER HUMAN
40	66.5	9.3	925	1	DBL HUMAN
41	66.5	9.3	1174	1	YJ11 YEAST
42	66	9.2	256	1	FRI3 VIGUN
43	66	9.2	389	1	CYB MYCVI
44	66	9.2	617	1	BFSI RAT
45	66	9.2	644	1	RNR_HELPY

ALIGNMENTS

RESULT 1
PROA_PSEPK
ID PROA_PSEPK STANDARD; PRT; 423 AA.
AC Q88DL4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Gamma-glutamyl phosphate reductase (GPR) [EC 1.2.1.41] (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase) (GSA dehydrogenase).
GN PROA OR PP4811.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinl C., Paulsen I.T., Dodson R.J., Hilbert H., Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M., Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J., Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I., Chris Lee P., Holtzapfel B., Scanlan D., Tran K., Moazzes A., Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S., Kiewitz C., Eisen J.A., Timmis J.A., Duesterhoeft A., Tuemmler B., Fraser C.M.;
RA "Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440.";
RT Environ. Microbiol. 4:799-808(2002).
RL CC
CC -!- FUNCTION: Catalyzes the NADPH dependent reduction of L-gamma-glutamyl 5-phosphate into L-glutamate 5-semialdehyde and phosphate. The product spontaneously undergoes cyclization to form 1-pyrroline-5-carboxylate.
CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate + NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
CC -!- PATHWAY: Proline biosynthesis; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: Belongs to the gamma-glutamyl phosphate reductase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE016792; AA070380.1; -.
DR TIGR; PP4811; -.
DR HAMAP; MF 00412; -; 1.
DR InterPro; IPR000965; Gglut_pp_reduct.
DR PROSITE; PS01223; PROA; 1.
KW Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 423 AA; 45417 MW; E2B9930ABE513A11 CRC64;

```
Query Match 11.4%; Score 81.5; DB 1; Length 423;
Best Local Similarity 20.5%; Pred. No. 1.2;
Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;

QY 12 LQYFSAGCKTPI--HLKDGVCALYNDEDAAVLEVPQHSQ-----50
Db 213 IERISDARVPVFKHL-DGICHY-----VSQHADLDKAWNVAFNKTYRYGIC 260
QY 51 -----SLLHCRITTEADPQTSITLYSMLLQNFENWAAWRC-----WLA- 89
Db 261 GAMETLLVDQVAER-----FLPEWARREVEKGVGLRGECRTQAIIISAKPATEADWHT 314
QY 90 -LDELHNVRLC--FQSSLEHDEASFSDIVSGTIEHAAVREYIAQLDESS 137
Db 315 YLDAILSRVVDGLNQLAIEHINHYGSHHTDSIISEHQGEARQFPAEVDSSAS 365

RESULT 2
AD29 HUMAN STANDARD; PRT; 820 AA.
AC Q9UKF5; Q9UHP1; Q9UKF3; Q9UKF4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAM 29 precursor (A disintegrin and metalloproteinase domain 29).
GN ADAM29.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA; BETA AND GAMMA).
RC TISSUE=Testis;
RX MEDLINE=99443746; PubMed=10512762;
RA Cerretti D.P., Dubose R.F., Black R.A., Nelson N.;
RT "Isolation of two novel metalloproteinase-disintegrin (ADAM) cDNAs
RT that show testis-specific gene expression."
RL Biochem. Biophys. Res. Commun. 263:810-815(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Testis;
RX MEDLINE=20112771; PubMed=10644455;
RA Xu R., Cai J., Xu T., Zhou W., Ying B., Deng K., Zhao S., Li C.;
RT "Molecular cloning and mapping of a novel ADAM gene (ADAM29) to human
RT chromosome 4."
RL Genomics 62:537-539(1999).
CC -!- FUNCTION: May be involved in spermatogenesis and fertilization.
CC Seems to be a non catalytic metalloprotease-like protein.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=Alpha;
CC IsoId=Q9UKF5-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q9UKF5-2; Sequence=VSP_005491;
CC Name=Gamma;
CC IsoId=Q9UKF5-3; Sequence=VSP_005492, VSP_005493;
CC -!- TISSUE SPECIFICITY: Expressed specifically in testes.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 disintegrin domain.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF171929; AAF03777.1; -.
DR EMBL; AF171930; AAF03778.1; -.
DR EMBL; AF171931; AAF03779.1; -.
DR EMBL; AF134708; AAF22163.1; -.
DR HSP; P18619; 1FVL.
DR MEROPS; M12.981; -.
DR Genew; HGNC:207; ADAM29.
DR MIM; 604778; -.
DR GO; GO:0005887; C: integral to plasma membrane; TAS.
DR GO; GO:0008237; F: metalloproteinase activity; TAS.
DR GO; GO:0007283; P: spermatogenesis; TAS.
DR GO; GO:0007283; P: spermatogenesis; TAS.
DR InterPro; IPR006586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR006025; Pept M Zn BS.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; P000289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
DR PROSITE; PS0214; DISINTEGRIN_2; 1.
DR PROSITE; PS00022; EGF_1; FALSE NEG.
DR PROSITE; PS01186; EGF_2; FALSE NEG.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
DR Signal; Glycoprotein; Transmembrane; EGF-like domain;
KW Alternative splicing; Repeat.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 193 BY SIMILARITY.
FT CHAIN 194 820 ADAM 29.
FT DOMAIN 194 674 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 675 695 POTENTIAL.
FT DOMAIN 696 820 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 194 390 METALLOPROTEASE-LIKE.
FT DOMAIN 397 483 DISINTEGRIN-LIKE.
FT DOMAIN 484 624 CYS-RICH.
FT DOMAIN 625 654 EGF-LIKE.
FT DOMAIN 659 675 POLY-LYS.
FT DOMAIN 739 819 9 X 9 AA APPROXIMATE REPEATS.
FT REPEAT 739 747 1.
FT REPEAT 748 756 2.
FT REPEAT 757 765 3.
FT REPEAT 766 774 4.
FT REPEAT 775 783 5.
FT REPEAT 784 792 6.
FT REPEAT 793 801 7.
FT REPEAT 802 810 8.
FT REPEAT 811 819 9.
FT DISULFID 307 384 BY SIMILARITY.
FT DISULFID 347 369 BY SIMILARITY.
FT DISULFID 455 468 POTENTIAL.
FT DISULFID 625 636 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT DISULFID 644 653 BY SIMILARITY.
FT CARBOHYD 217 217 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 538 538 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARSPLIC 750 803 Missing (in isoform Beta).
FT VARSPLIC 763 787 /FTID=VSP_005491.
FT VARSPLIC 788 820 VMPSSHQPLTSPSQPPVMPQSH -> QNLFLEFSISD
CVLNFRLYLOAT (in isoform Gamma).
FT /FTID=VSP_005492.
FT Missing (in isoform Gamma).
FT /FTID=VSP_005493.
```

```

FT CONFLICT 196 H -> Y (IN REF. 2).
FT CONFLICT 744 P -> H (IN REF. 2).
FT CONFLICT 748 S -> Y (IN REF. 2).
FT CONFLICT 753 Q -> R (IN REF. 1); AAF03777).
FT CONFLICT 764 M -> T (IN REF. 2).
FT CONFLICT 769 HPQLT -> QPRVM (IN REF. 2).
SQ SEQUENCE 820 AA; 92753 MW; 1F54E9F8128E4C27 CRC64;

Query Match 11.3%; Score 80.5; DB 1; Length 820;
Best Local Similarity 22.3%; Pred. No. 3.1;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPIHLK-----DGVCALYNEQDEAAVLEVPQHSDDLHCHRIIEADPQTSITLYSMILLQNFEMAA 82
Db 60 GQKHIIHKVKKLSKHLPTTYTDQGAILEDQPFVQNNCYH-GYVEGDPFESLSLST 118
QY 70 ----YSMLLQNFEMAAAMRGWLALDELHNVR-LCFQOSLEHL-----DEASRSDIVSG 118
Db 119 CFGGFOGILQIN-----DFAYEIKPLAFSTFHELVYKMDSEKQPFSTRMSG 165
QY 119 FIEHAAVREYIAQDESS 137
Db 166 FMQNEITCRMEFEIDNST 184

RESULT 3
PLSB_MYCTU
ID PLSB_MYCTU STANDARD; PRT; 789 AA.
AC 053207;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycerol-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR PLSB2 OR RV2482C OR MT2555 OR MTV008.38C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Taylor K.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J.A., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
CC acyl-sn-glycerol 3-phosphate.
CC -!- PATHWAY: De novo phospholipid biosynthesis; first step. May also
CC function in the regulation of membrane biogenesis.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the GPAT / DAPAT family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021246; CAAL16059.1; -.
DR EMBL; AE007092; AAK46859.1; -.
DR PIR; A70868; A70868.
DR TIGR; MT2555; -.
DR TubercuList; RV2482C; -.
DR HAMAP; MF_00393; -.
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; pIsC; 1.
KW Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
KW Complete proteome.
FT CONFLICT 180 R -> Q (IN REF. 2).
FT CONFLICT 778 C -> R (IN REF. 2).
SQ SEQUENCE 789 AA; 88314 MW; B78D7D8F9296EA22 CRC64;

Query Match 11.0%; Score 79; DB 1; Length 789;
Best Local Similarity 24.3%; Pred. No. 4.2;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

QY 23 IHLKDGVCALYNEQDEAAVLEVPQHSDDLHCHRIIEADPQTSITLYSMILLQNFEMAA 82
Db 445 VSMRQYLGAPHGELTQDPAKELALQKMSFEVAVRILQATPVTATGIVSALL-----LT 498
QY 83 MRGWLALDELHNVRILCPQOSLEHL-----EASFSDIVSGF----- 119
Db 499 TRGTALTLDQLHHT---LQDSLVDYLRQSPVSTALRLRSRGVRAAADALSNGHPVTR 555
QY 120 IEHAAVREYIAQDESSAA 139
Db 556 VDSGREGPVWYIAPDDEHAAA 575

RESULT 4
PLSB_MYCLE
ID PLSB_MYCLE STANDARD; PRT; 775 AA.
AC Q9X7B0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycerol-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR ML1246 OR MLCB1610.07.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
CC acyl-sn-glycerol 3-phosphate.
CC -!- PATHWAY: De novo phospholipid biosynthesis; first step. May also
CC function in the regulation of membrane biogenesis.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the GPAT / DAPAT family.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL049913; CAB43153.1; -
CC DR EMBL; AL583921; CAC31627.1; -
CC DR PIR; T45238; T45238.
CC DR Leproma; ML1246; -
CC DR HAMAP; MF_00393; -; 1.
CC DR InterPro; IPR002123; Acyltransferase.
CC DR Pfam; PF01553; Acyltransferase; 1.
CC DR SMART; SM00563; Plac; 1.
CC DR Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
CC KW Complete proteome.
CC SEQUENCE 775 AA; 87363 MW; 03DD77C778293CDF CRC64;
CC -----
CC Query Match 10.8%; Score 77.5; DB 1; Length 775;
CC Best Local Similarity 25.8%; Pred. No. 5.9;
CC Matches 34; Conservative 16; Mismatches 33; Indels 49; Gaps 5;
CC -----
CC QY 11 ELOYFSGAGKTPHLKDGVCALYN-----EQ 36
CC Db 397 YAAVARGKEKTP-----EGVWLKSFKAQGERNYGKIYVFPPEAVSRQVILGAPHGALVQ 452
CC QY 37 DEEAVALVEQPHSDLSLHLCRIIRADPQTSITLYSMLLQLNFMMAVRGCGWLALDELHNV 96
CC Db 453 DQDAKRLALQKMSPEVA--WRILCATPTATALVSALL-----LITRGVALTLDQLLHT 504
CC QY 97 RLCPQOQSLEHL 108
CC Db 505 ---LQESLDYLE 513
CC -----
CC RESULT 5
CC SSN6_YEAST STANDARD; PRT; 966 AA.
CC ID SSN6_YEAST
CC AC P14922;
CC DT 01-APR-1990 (Rel. 14, Last sequence update)
CC DT 01-APR-1990 (Rel. 14, Last sequence update)
CC DT 01-FEB-1995 (Rel. 31, Last annotation update)
CC DE Glucose repression mediator protein.
CC GN SSN6 OR CYC8 OR YBR112C OR YBR0908.
CC OS Saccharomyces cerevisiae (Baker's yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CC OX NCBI_TaxID=4932;
CC RN [1]
CC SEQUENCE FROM N.A.
CC RP MEDLINE=89211964; PubMed=2854095;
CC RA Trumbly R.J.;
CC RT "Cloning and characterization of the CYC8 gene mediating glucose
CC repression in yeast";
CC RL Gene 73:97-111(1988).
CC RN [2]
CC SEQUENCE FROM N.A.
CC RP MEDLINE=89065502; PubMed=3316983;
CC RA Schultz J., Carlson M.;
CC RT "Molecular analysis of SSN6, a gene functionally related to the SNF1
CC protein kinase of Saccharomyces cerevisiae";
CC RL Mol. Cell. Biol. 7:3637-3645(1987).
CC RN [3]
CC SEQUENCE FROM N.A.
CC RP STRAIN=S288C;
CC RA Mannhaupt G., Stucka R., Ehle S., Vetter I., Feldmann H.;
CC RT "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
CC excision repair gene RAD16 located in this region belongs to a novel
CC group of double-finger proteins";

Yeast 8:397-408(1992).
[4]
RL TPR REPEATS.
RX MEDLINE=90124639; PubMed=2404612;
RA Sikorski R.S., Boguski M.S., Goebel M., Hieter P.A.;
RT "A repeating amino acid motif in CDC23 defines a family of proteins
RT and a new relationship among genes required for mitosis and RNA
RT synthesis";
RL Cell 60:307-317(1990).
CC -1- FUNCTION: IT IS INVOLVED IN REPRESSION BY AL-ALPHA2 AND ALPHA2 AND
CC IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 10 TPR repeats.
CC -1- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M23440; AAA34545.1; -
CC DR EMBL; M17826; AAA35103.1; -
CC DR EMBL; X66247; CAA46973.1; -
CC DR EMBL; X78993; CAA5615.1; -
CC DR EMBL; Z35981; CAA85069.1; -
CC DR PIR; S25365; S25365.
CC DR Germonline; 138655; -
CC DR TRANSFAC; T03687; -
CC DR SGD; S0000316; CYC8.
CC DR GO; GO:0005634; C:nucleus; IPI.
CC DR GO; GO:0016565; F:general transcription co-activator activity; IDA.
CC DR GO; GO:0003713; F:transcription co-activator activity; IDA.
CC DR GO; GO:0016481; P:negative regulation of transcription; IDA.
CC DR InterPro; IPR008941; TPR-like.
CC DR InterPro; IPR001440; TPR.
CC DR Pfam; PF00515; TPR; 10.
CC DR SMART; SM00028; TPR; 9.
CC DR KW Transcription regulation; Repressor; Repeat; TPR repeat;
CC Nucleic acid binding protein.
CC FT DOMAIN 15 30 POLY-GLN.
CC FT REPEAT 46 79
CC FT REPEAT 80 113
CC FT REPEAT 114 147
CC FT REPEAT 150 183
CC FT REPEAT 187 220
CC FT REPEAT 224 257
CC FT REPEAT 258 291
CC FT REPEAT 296 329
CC FT REPEAT 330 363
CC FT REPEAT 364 398
CC FT DOMAIN 493 556 30 X 2 AA TANDEM REPEATS OF Q-A.
CC FT DOMAIN 557 587 POLY-GLN.
CC FT CONFLICT 547 547 K -> Q (IN REF. 3).
CC SEQUENCE 966 AA; 107202 MW; 84B509CF3208C5C0 CRC64;
CC -----
CC Query Match 10.6%; Score 75.5; DB 1; Length 966;
CC Best Local Similarity 26.2%; Pred. No. 12;
CC Matches 22; Conservative 15; Mismatches 34; Indels 13; Gaps 2;
CC -----
CC QY 64 QTSITLYSMLLQLNFM-----AAMRGCGWLALDELHNVRLCPQOQSLEHLDEAFSDIVSGF 119
CC Db 97 QRAAEYERALLVNPFLSDVWATLGHCHYLMLDLDLQRAYNAVYQQAALYHLSNPNVFKLWGI 156
CC QY 120 -----IEHAAEVRREYIAQLD 134
CC Db 157 GILYDRYCSLDYAEAEFAKVLIED 180

FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 338 338 A -> G (IN REF. 1).
 FT CONFLICT 581 616 ORDRATDVEDREKMLKPVVDLQAYEDPAQGALDFT ->
 HVTAPPMWERTSCABALGCTSRHRTLHREFTWTLPGWSN
 FT FFS (IN REF. 1).
 SQ SEQUENCE 976 AA; 108080 MW; 50512CCF13F03808 CRC64;
 Query Match 10.3%; Score 73.5; DB 1; Length 976;
 Best Local Similarity 27.5%; Pred. No. 19;
 Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6;
 QY 7 RVERPLOVFSACKT-----PIHLKGVGCALYNEQDEAAVLEVPQHSDSLHLHCRITE 60
 DB 94 RVHVELQFTVRDCKSPFGAGPLGCKETFNLLYMESDQDVGI---QLRRPLFKQVTTVA 149
 QY 61 ADPQTSI-TLYSMLQLNFEMAAM-----RGCLWALDE-----LHNVRLCFQOSLEHLD 108
 DB 150 ADQSFTRDLASGVKLVNVERCSGLRGLTRGLYLAFNPGACVALSVRVFYQRCPETLN 209
 QY 109 E-ASFSDIVSG 118
 DB 210 GLAQFPDTLPG 220
 RESULT 9
 ID PROA PSESM STANDARD; PRT; 421 AA.
 AC Q87VU6;
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DE Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-
 semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde
 dehydrogenase) (GSA dehydrogenase).
 GN PROA OR PSPT04823.
 OS Pseudomonas syringae (pv. tomato).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DC3000;
 RX MEDLINE=22834015; PubMed=12928499;
 RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
 RA Gwinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
 RA Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
 RA Nelson W.C., Davidson T., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
 RA Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
 RA Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P.,
 RA Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
 RA White O., Fraser C.M., Collier A.,
 RA "The complete genome sequence of the Arabidopsis and tomato pathogen
 RT Pseudomonas syringae pv. tomato DC3000.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
 CC -!- FUNCTION: Catalyzes the NADPH dependent reduction of L-gamma-
 glutamyl 5-phosphate into L-glutamate 5-semialdehyde and
 CC phosphate. The product spontaneously undergoes cyclization to form
 CC 1-pyrroline-5-carboxylate.
 CC -!- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
 CC NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
 CC -!- PATHWAY: Proline biosynthesis; second step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the gamma-glutamyl phosphate reductase
 CC family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB016873; AAO58258.1; --
 DR TIGR; FSP04829; -- 1.
 DR HAMAP; MF_00412; -- 1.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR InterPro; IPR000965; Gglut_pp_reduct.
 DR Pfam; PF00171; aldedh; 1.
 DR PROSITE; PS01223; PROA; 1.
 DR Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.
 KW SEQUENCE 421 AA; 45003 MW; 8FFC78CF3E2A3483 CRC64;
 SQ
 Query Match 10.1%; Score 72.5; DB 1; Length 421;
 Best Local Similarity 22.0%; Pred. No. 9.1;
 Matches 38; Conservative 29; Mismatches 39; Indels 67; Gaps 11;
 QY 12 IQYFSGACKTPI--HLKGVGCALYNEQDEAAVLEVPQHSD----- 50
 DB 213 IERSVSDARVPVVKHL-DGICHVY-----VSAHADLPKAKTAFNAKTYRGIC 260
 QY 51 ---SLLHCRITADPQTSITLYSMLQLNFEMAAMRG-----WLA-- 89
 DB 261 GAMETLLVDQTIAAD-----FLPMAAQFKEKGVLCERTRELIDYMPATEDDMHTEY 315
 QY 90 LDELHNVRLC--FQOSLEHLD--ASFSD-IVSGFIEHAAEVREYIAQLDESS 137
 DB 316 LAAILSVSVGLDEAIEHINHYGSHSDAIVS---DHQSQIRRRMAEVDSSS 365
 RESULT 10
 ID OGTL HUMAN STANDARD; PRT; 1036 AA.
 AC Q15294; Q96CC1;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase 110
 DE kDa subunit [EC 2.4.1.-] (O-GlcNAc transferase p110 subunit).
 GN OGT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND SEQUENCE OF 217-226 AND 945-961.
 RC TISSUE=Liver;
 RX MEDLINE=97238870; PubMed=9083068;
 RA Lubas W.A., Frank D.W., Krause M., Hanover J.A.;
 RA "O-linked GLCNAC transferase is a conserved nucleocytoplasmic protein
 RT containing tetratricopeptide repeats.";
 RL J. Biol. Chem. 272:9316-9324(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=21635556; PubMed=11773972;
 RA Nolte D., Muller U.;
 RT "Human O-GlcNAc transferase (OGT): genomic structure, analysis of
 RL splice variants, fine mapping in Xq31.1.";
 RN Mamm. Genome 13:62-64(2002).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.M., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udeshi T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- FUNCTION: Addition of nucleotide-activated sugars directly onto
 the polypeptide through O-glycosidic linkage with the hydroxyl of
 serine or threonine.
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + peptide = UDP +
 N-acetyl-beta-D-glucosaminyl-peptide.
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBUNIT: HETEROTRIMER OF TWO 11 kDa AND ONE 70 kDa SUBUNITS. IT IS
 NOT KNOWN IF THE 70 kDa SUBUNIT IS ENCODED BY A SEPARATE GENE OR
 IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN
 ALTERNATIVE INITIATION OF THE 110 kDa SUBUNIT (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (Possible).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=O15294-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=O15294-2; Sequence=VSP 006553;
 CC -!- TISSUE SPECIFICITY: Highly expressed in pancreas and to a lesser
 extent in skeletal muscle, heart, brain and placenta. Present in
 trace amounts in lung and liver.
 CC -!- SIMILARITY: Contains 12 TPR repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

 EMBL; U77413; AAB63466.1; -;
 DR EMBL; AJ315767; CAC86127.1; -;
 DR EMBL; AJ315767; CAC86129.1; -;
 DR EMBL; BC014434; AAH14434.1; -;
 DR Genbank; HGNC:8127; OGT.
 DR MIM; 300255; -;
 DR GO; GO:0005829; C:cytosol; TAS.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0008375; F:acetylglucosaminyltransferase activity; TAS.
 DR GO; GO:0005515; F:protein binding; TAS.
 DR GO; GO:0006493; P:O-linked glycosylation; TAS.
 DR GO; GO:0007584; P:response to nutrients; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 12.
 DR SMART; SM00028; TPR; 11.
 KW Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;
 KW Alternative splicing
 FT REPEAT 79 112 TPR 1.
 FT REPEAT 113 146 TPR 2.
 FT REPEAT 147 180 TPR 3.
 FT REPEAT 181 214 TPR 4.
 FT REPEAT 215 248 TPR 5.
 FT REPEAT 249 282 TPR 6.
 FT REPEAT 283 316 TPR 7.
 FT REPEAT 317 350 TPR 8.
 FT REPEAT 351 384 TPR 9.
 FT REPEAT 385 418 TPR 10.
 FT REPEAT 419 452 TPR 11.
 FT REPEAT 453 463 TPR 12 (INCOMPLETE).
 FT DOMAIN 477 493 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT VARSPLIC 1 166 MASSGVNADSTGLAEHLAREYQAGFEAEHCHMLWROE
 PONTGVLLLSLSIHFCRRDKRSAAHFTSLIAIKQNPILAEAY

FT SNLGNVYKRGQIOEAIEHYRHALRLKPDIDGYNLAAAL
 FT VAAGDMGAVQAVSALQYNPDLYCVRSDLGNLILKALGRLE
 FT EA -> MLOGHFWLVRGIMISPSPPPNLFFPLQIFPF
 FT PFTSPSHLSLTPP (in isoform 2).
 FT /FTId=VSP_006553.
 SQ SEQUENCE 1036 AA; 115705 MW; C3BD67340925A2C2 CRC64;
 Query Match 10.0%; Score 71.5; DB 1; Length 1036;
 Best Local Similarity 24.4%; Pred. No. 32;
 Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;
 Qy 24 HLKDGVCALYNEQDEEAIVLEV-----POHSDSL--LHCHRIIADPQTSITLY 70
 Db 281 HFFDAYCNLANALKEGSAEADCVNTALRLCTHADSLLNLANIKREQGNIEEVRLY 340
 Qy 71 SMILLQNFMAAMRGCGWLDLHNVRLCFQSLHLEDEA-----SFSVDVSGFIEHAAE 125
 Db 341 RKALEVEPEPAHAHS---NLASVLQOQKGLQEALMHWYKEAIRISPTFADAYSNGMTLKE 397
 Qy 126 VRE 128
 Db 398 MQD 400
 RESULT 11
 OGT1 RAT
 ID -OGT1 RAT STANDARD; PRT; 1036 AA.
 AC P56558;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase 110
 kDa subunit (EC 2.4.1.-) (O-GlcNAc transferase p110 subunit).
 GN OGT.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RX STRAIN=Sprague-Dawley; Tissue=Liver;
 RX MEDLINE=97238869; PubMed=9083067;
 RA Kreppel L.K., Blomberg M.A., Hart G.W.;
 RT "Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and
 characterization of a unique O-GlcNAc transferase with multiple
 tetratricopeptide repeats.";
 RL J. Biol. Chem. 272:9308-9315 (1997).
 CC -!- FUNCTION: Addition of nucleotide-activated sugars directly onto
 the polypeptide through O-glycosidic linkage with the hydroxyl of
 serine or threonine.
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + peptide = UDP +
 N-acetyl-beta-D-glucosaminyl-peptide.
 CC -!- ENZYME REGULATION: BY TYROSINE PHOSPHORYLATION AND O-GLCNAc
 MODIFICATIONS.
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBUNIT: HETEROTRIMER OF TWO 110 kDa AND ONE 78 kDa SUBUNITS. IT
 IS NOT KNOWN IF THE 78 kDa SUBUNIT IS ENCODED BY A SEPARATE GENE
 OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN
 ALTERNATIVE INITIATION OF THE 110 kDa SUBUNIT.
 CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (Possible).
 CC -!- TISSUE SPECIFICITY: Appears to be present in all tissues examined
 except kidney.
 CC -!- PTM: AUTOMODIFIED BY O-GLYCOSYLATION WITH O-GLCNAc.
 CC -!- SIMILARITY: Contains 12 TPR repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

```

DR EMBL; U76557; AAC53121.1; -
DR PIR; T31673; T31673.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 12.
DR SMART; SM00028; TPR; 11.
KW Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;
KW Phosphorylation; Glycoprotein.
FT REPEAT 79 112 TPR 1.
FT REPEAT 113 146 TPR 2.
FT REPEAT 147 180 TPR 3.
FT REPEAT 181 214 TPR 4.
FT REPEAT 215 248 TPR 5.
FT REPEAT 249 282 TPR 6.
FT REPEAT 283 316 TPR 7.
FT REPEAT 317 350 TPR 8.
FT REPEAT 351 384 TPR 9.
FT REPEAT 385 418 TPR 10.
FT REPEAT 419 452 TPR 11.
FT REPEAT 453 483 TPR 12 (INCOMPLETE).
FT DOMAIN 478 493 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 979 979 PHOSPHORYLATION (PROBABLE).
SQ SEQUENCE 1036 AA; 115605 MW; 3F057CABDD019BD6 CRC64;

Query Match
Best Local Similarity 10.0%; Score 71.5; DB 1; Length 1036;
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

QY 24 HLKDGVCALYNQDEEAAVLEV-----POHSDSL--LLHCRILEADPOTSITLY 70
Db 281 HFPDAYCNLANALKKGVAEEDCVNTRALRCLPTHADSLNNLANIKRQGNIEAAVRLY 340

QY 71 SMLLGINFEMAAMRGCGWLALDELHNVRLCFQOSLEHLDEA----SFSDIVSGFIEHAAE 125
Db 341 RKALEVFFFEAAHS---NLASVLQOQGLQELAHMXYKEAIRISPTFADAYSNMGNLTKE 397

QY 126 VRE 128
Db 398 MQD 400

RESULT 12
NEL2 MOUSE
ID NEL2_MOUSE STANDARD; PRT; 816 AA.
AC Q61220;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein kinase C-binding protein NELL2 precursor (NEL-like protein 2)
DE (MEL91 protein).
GN NELL2 OR MEL91.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1] _TaxID=10090;
[2]
SEQUENCE FROM N.A.
RA Elkins D.A., Rossi J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RP TSP N-TERMINAL DOMAIN.
RX MEDLINE=98153258; PubMed=9480764;
RA Beckmann G., Hanke J., Bork P., Reich J.;
RT "Merging extracellular domains: fold prediction for laminin G-like
RT and amino-terminal thrombospondin-like modules based on homology to
RT pentraxins.";
RL J. Mol. Biol. 275:725-730(1998).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -!- SIMILARITY: Contains 5 VWFC domains.
CC -!- SIMILARITY: Contains 6 EGF-like domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC EMBL; U59230; AAB02924.1; ALT_INIT.
CC HSSP; P00740; IEDM.
CC MGD; MGI:1858510; Nell2.
CC InterPro; IPR00152; Asx_hydroxyl_S.
CC InterPro; IPR008985; ConA_like_lec_gl.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; VWFC_C.
CC Pfam; PF00008; EGF; 4.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00093; VWFC; 2.
CC SMART; SM00179; EGF_CA; 3.
CC SMART; SM00282; LamG; 1.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00214; VWFC; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; 3.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 3.
CC PROSITE; PS00026; EGF_3; 6.
CC PROSITE; PS01187; EGF_CA; 3.
CC PROSITE; PS01208; VWFC_1; 2.
CC PROSITE; PS01184; VWFC_2; 2.
KW Glycoprotein; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 816 PROTEIN KINASE C-BINDING PROTEIN NELL2.
FT DOMAIN 272 331 TSP N-TERMINAL.
FT DOMAIN 332 396 VWFC 1.
FT DOMAIN 397 439 EGF-LIKE 1.
FT DOMAIN 440 481 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 482 522 EGF-LIKE 3.
FT DOMAIN 521 553 EGF-LIKE 4.
FT DOMAIN 555 601 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 602 637 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 638 693 VWFC 3.
FT DOMAIN 698 756 VWFC 4.
FT DOMAIN 758 813 VWFC 5.
FT DISULFID 401 413 BY SIMILARITY.
FT DISULFID 407 422 BY SIMILARITY.
FT DISULFID 424 438 BY SIMILARITY.
FT DISULFID 444 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 480 BY SIMILARITY.
FT DISULFID 486 499 BY SIMILARITY.
FT DISULFID 493 508 BY SIMILARITY.
FT DISULFID 510 521 BY SIMILARITY.
FT DISULFID 525 535 BY SIMILARITY.
FT DISULFID 529 541 BY SIMILARITY.
FT DISULFID 543 552 BY SIMILARITY.
FT DISULFID 559 572 BY SIMILARITY.
FT DISULFID 566 581 BY SIMILARITY.
FT DISULFID 583 600 BY SIMILARITY.
FT DISULFID 606 619 BY SIMILARITY.
FT DISULFID 613 628 BY SIMILARITY.
FT DISULFID 630 636 BY SIMILARITY.
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 816 AA; 91163 MW; 5BDD0A946F87E74D CRC64;
9.9%; Score 71; DB 1; Length 816;
Query Match

```

Best Local Similarity 24.6%; Pred. No. 28;
Matches 33; Conservative 12; Mismatches 55; Indels 34; Gaps 4;
QY 3 SSQORVERFLQ-----YFAGCKPIHLKDGVCALYNQDEEAAVLEVPQHSDSLH 55
Db 69 ASATATREFLQKLNKHEFTILVTLKQHLNLSGVILSHLDHRYLESNGHRNEIRLH 128
QY 56 CRRIEADPQTSITLYSMLLQNFEMAAAMRGWGLALDELHNVRLCFQOS--LEHLD----- 108
Db 129 YRSGTHRPHTVEVPYI-----LADAKWHKLSAFSASHLILHIDCKNIY 172
QY 109 -----EASFSDIVSG 118
Db 173 ERVVEPFTDLALG 186
RESULT 13
DOC8 HUMAN STANDARD; PRT; 1799 AA.
AC QNF50; Q8TEP1; Q8WUY2; Q9BYJ5; Q9H1Q2; Q9H1Q3; Q9H308; Q9H7P2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dedicator of cytokinesis protein 8 (fragment).
GN DOCK8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Spleen;
RA Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;
RT "the nucleotide sequence of a long cDNA clone isolated from human
spleen.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Laird G., Skuce C.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 1194-1799 FROM N.A.
RP TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.T., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]
RN SEQUENCE OF 1475-1799 FROM N.A.
RX MEDLINE=20195627; PubMed=10729223;
RA Otterleghi C., Veitia R., Quintana-Murci L., Torchara D., Scapoli L.,
RA Souleyreau-Therville N., Beckmann J., Fellous M., McElreavey K.;
RT "The region on 9p associated with 46,XY sex reversal contains several
transcripts expressed in the urogenital system and a novel
doublesex-related domain.";
RL Genomics 64:170-178 (2000).

QY 24 HKDGVCMAL--YFQDEEAAVLEVPQHSDSL--LHCRIEADPQTSITLYS 71
Db 1269 HLRSRLTILAYSEEDTAMQTPPTQVEILLNLSILYDTVKMREFQDPE----- 1321
QY 72 MLLQLNPFMAA-----MRGCLLA-LDELHNVRLCFQOSLEHLDASFSIVSGFIEHA 123
Db 1322 MMDLMYRIAKSYQASPDRLTLWLNMAEKKTKKCYTEAMCL-----VHAA 1369
QY 124 AEVREYIAQLDESS 137
Db 1370 ALVAEYLSMLEDHS 1383
Query Match 9.9%; Score 70.5; DB 1; Length 1799;
Best Local Similarity 21.6%; Pred. No. 77;
Matches 29; Conservative 24; Mismatches 42; Indels 39; Gaps 6;
QY 24 HKDGVCMAL--YFQDEEAAVLEVPQHSDSL--LHCRIEADPQTSITLYS 71
Db 1269 HLRSRLTILAYSEEDTAMQTPPTQVEILLNLSILYDTVKMREFQDPE----- 1321
QY 72 MLLQLNPFMAA-----MRGCLLA-LDELHNVRLCFQOSLEHLDASFSIVSGFIEHA 123
Db 1322 MMDLMYRIAKSYQASPDRLTLWLNMAEKKTKKCYTEAMCL-----VHAA 1369
QY 124 AEVREYIAQLDESS 137
Db 1370 ALVAEYLSMLEDHS 1383
[5]
RN SEQUENCE OF 1595-1799 FROM N.A.
RP TISSUE=Lymph node;
RA Koehrer K., Beyer A., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[6]
RN NOMENCLATURE.
RX MEDLINE=22319137; PubMed=12432077;
RA Cote J.-F., Vuori K.;
RT "Identification of an evolutionarily conserved superfamily of DOCK180-
related proteins with guanine nucleotide exchange activity.";
J. Cell Sci. 115:4901-4913 (2002).
CC -!- FUNCTION: Potential guanine nucleotide exchange factor (GEF). GEF
proteins activate some small GTPases by exchanging bound GDP for
free GTP (By similarity).
CC -!- DOMAIN: The DHR-2 domain may mediate some GEF activity (By
similarity).
CC -!- SIMILARITY: Belongs to the DOCK family.
CC -!- SIMILARITY: Contains 1 DHR-1 (CZH-1) domain.
CC -!- SIMILARITY: Contains 1 DHR-2 (CZH-2) domain.
CC -!- CAUTION: Ref.1 (BAB84907) sequence differs from that shown due to
a frameshift in position 1656.
CC -!- CAUTION: Ref.2 (CAC22149) sequence differs from that shown due to
erroneous gene model prediction.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to
frameshifts in positions 1767 and 1782.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AK024436; BAB15726.1; -
EMBL; AK074081; BAB84907.1; ALT_FRAME.
EMBL; AK090429; BAC03410.1; -
EMBL; AL158832; -; NOT ANNOTATED_CDS.
EMBL; AL161725; CAC22148.1; -
EMBL; AL161725; CAC22149.1; ALT_SEQ.
EMBL; BC019102; AAH19102.1; ALT_INIT.
EMBL; AF194407; AAC42221.1; ALT_FRAME.
EMBL; AL583913; CAC29497.1; -
Genew; HGNC:19191; DOCK8.
KW Guanine-nucleotide releasing factor.
FT NON_TER 1 DHR-1.
FT DOMAIN 260 502 DHR-2.
FT DOMAIN 1234 1755 SG -> C (IN REF. 2).
FT CONFLICT 568 569 V -> F (IN REF. 1; BAB84907).
FT CONFLICT 1451 1451 V -> F (IN REF. 1; BAB84907).
FT CONFLICT 1455 1455 A -> P (IN REF. 4).
FT CONFLICT 1670 1670 E -> K (IN REF. 4).
FT CONFLICT 1729 1729 L -> F (IN REF. 4).
FT CONFLICT 1746 1746 L -> F (IN REF. 4).
SQ SEQUENCE 1799 AA; 204390 MW; 29FECA062A0E16 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:13:22 ; Search time 6.18715 Seconds
(without alignments)
7088.403 Million cell updates/sec

Title: US-09-596-784-4
Perfect score: 715
Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_todent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	2	O54621 erwinia amy
2	455	63.6	124	2	Q9KH43
3	442	61.8	139	2	Q9FCV6
4	273.5	38.3	129	2	O66102 pseudomonas
5	273.5	38.3	131	16	Q9JF37
6	86.5	12.1	142	16	Q88BP9 pseudomonas
7	81.5	11.4	423	16	Q88DL4 pseudomonas
8	81	11.3	186	16	Q92RQ7
9	81	11.3	723	17	Q8ZW74
10	80.5	11.3	368	10	Q40879
11	79.5	11.1	1021	16	Q40879 pelargonium
12	79	11.0	708	16	Q81R76
13	79	11.0	789	16	Q7TVH5
14	78.5	11.0	830	16	Q9P332
15	78	10.9	155	2	Q8RP06
16	78	10.9	830	16	Q87E45 xylella fas

17	77.5	10.8	300	16	Q8F9X7 leptospira
18	77	10.8	130	16	Q87W45 pseudomonas
19	76.5	10.7	748	4	Q86VZ7 homo sapien
20	76	10.6	371	10	Q8RVM0 coffea cane
21	75.5	10.6	212	16	Q8A7K9 bacteroides
22	75.5	10.6	402	5	O45260 caenorhabdi
23	75.5	10.6	468	16	Q8PN02 xanthomonas
24	75.5	10.6	1588	11	Q9ESK9 mus musculu
25	75	10.5	317	2	Q9FD53 streptococc
26	75	10.5	317	16	Q97S10
27	75	10.5	923	13	Q7ZVM0 brachydanio
28	75	10.5	2077	5	Q9XIM2 drosophila
29	74.5	10.4	549	5	Q86SC9 ciona intes
30	74.5	10.4	1051	5	Q9VA99 drosophila
31	74.5	10.4	1109	5	Q8IM19
32	74	10.3	304	16	Q8PBE7 xanthomonas
33	74	10.3	443	2	Q93A56 pseudomonas
34	73.5	10.3	751	11	Q7TMQ5 mus musculu
35	73.5	10.3	892	4	Q9HCF0 homo sapien
36	73	10.2	578	11	Q62693 rattus norv
37	73	10.2	643	10	Q9FTE5 oryza sativ
38	73	10.2	678	16	Q87MS2 vibrio para
39	73	10.2	1708	16	Q8YZI2 anabaena sp
40	72.5	10.1	131	16	Q8YV78 pseudomonas
41	72.5	10.1	421	16	Q87VV6 mus musculu
42	72.5	10.1	678	11	Q8BRV9 mus musculu
43	72.5	10.1	1467	11	Q8CHH8 mus musculu
44	72.5	10.1	1958	2	Q93CG6 photobacter
45	72.5	10.1	2183	12	Q9E588 subacute sc

ALIGNMENTS

RESULT 1
O54621
ID O54621 PRELIMINARY; PRT; 139 AA.
AC O54621;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE DSPP.
GN DSPP OR DSPB.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EA321;
RX MEDLINE=98115919; PubMed=9448330;
RA Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA Conlin A.K., Collmer A., Beer S.V.;
RT "Homology and functional similarity of an hrp-linked pathogenicity
locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
Pseudomonas syringae pathovar tomato.";
RT Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EA321;
RX Bogdanove A.J., Kim J.F., Wei Z.-M., Kolchinsky P., Beer S.V.;
RT Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CFBP1430;
RX MEDLINE=98086111; PubMed=9426142;
RA Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;
RT "DspA, an essential pathogenicity factor of Erwinia amylovora showing
homology with AvrE of Pseudomonas syringae, is secreted via the Hrp
secretion pathway in a DspB-dependent way.";
RT Mol. Microbiol. 26:1057-1069(1997).
DR EMBL; U97504; AAC04851.1; -;
DR EMBL; Y13831; CAA74157.1; -;

DR PIR: T18449; T18449. 60F23F88B1BFAE93 CRC64;
SQ SEQUENCE 139 AA; 15723 MW; 60F23F88B1BFAE93 CRC64;
Query Match 100.0%; Score 715; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 1.2e-68;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTSSQQRVERFLOYSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
Db 1 MTSSQQRVERFLOYSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
QY 61 ADPQTSITLYSMLLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
Db 61 ADPQTSITLYSMLLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
QY 121 EHAAEVREYIAQLDESSAA 139
Db 121 EHAAEVREYIAQLDESSAA 139
RESULT 2
Q9KH43 PRELIMINARY; PRT; 124 AA.
AC Q9KH43;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE DsPF.
GN DsPF.
OS Pantoea agglomerans (pv. gypsophilae) (Erwinia herbicola).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=48984;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21171042; PubMed=11277443;
RA Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT "Genetic organization of the hrp gene cluster and dspAE/BF operon in
Erwinia herbicola pv. gypsophilae.";
RL Mol. Plant Microbe Interact. 14:431-436 (2001).
DR EMBL; AF271717; AAF76344.1; -;
SQ SEQUENCE 124 AA; 13873 MW; 9F0266FCB78CCEB0 CRC64;
Query Match 63.6%; Score 455; DB 2; Length 124;
Best Local Similarity 67.5%; Pred. No. 7.3e-41;
Matches 81; Conservative 18; Mismatches 21; Indels 0; Gaps 0;
QY 1 MTSSQQRVERFLOYSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
Db 1 MTPSQQRVDAFLQHFASCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
QY 61 ADPQTSITLYSMLLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
Db 61 TAPHSALNFYALLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
RESULT 3
Q9FCY6 PRELIMINARY; PRT; 139 AA.
AC Q9FCY6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE WTSF.
GN WTSF.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=66271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS104;
RA Frederick R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,

RA Manulis S., Coplin D.L.;
RT "Genetic organization of the Pantoea stewartii subsp. stewartii hrp
gene cluster and sequence analysis of the hrpA, hrpC, hrpN and wtsE
operons";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; A2282857; AG01468.1; -. 6290EAE8003D7A9 CRC64;
SQ SEQUENCE 139 AA; 15670 MW; 6290EAE8003D7A9 CRC64;
Query Match 61.8%; Score 442; DB 2; Length 139;
Best Local Similarity 58.3%; Pred. No. 2e-39;
Matches 81; Conservative 21; Mismatches 37; Indels 0; Gaps 0;
QY 1 MTSSQQRVERFLOYSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
Db 1 MTPSQKKEALLQQAATSCNTTLHLKEGVCALYNTQGEAAVLEVPQSDCLLHCRRIE 60
QY 61 ADPQTSITLYSMLLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
Db 61 AEHQAGQNFYALLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
QY 121 EHAAEVREYIAQLDESSAA 139
Db 121 TQAEVREFITHTEHDVA 139
RESULT 4
O66102 PRELIMINARY; PRT; 129 AA.
ID O66102
AC O66102; 1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Avirulence protein.
GN AVRF.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98115919; PubMed=9448330;
RA Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
Conlin A.K., Collmer A., Beer S.V.;
RT "Homology and functional similarity of an hrp-linked pathogenicity
locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
Pseudomonas syringae pathovar tomato.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330 (1998).
DR EMBL; U97505; AAC06135.1; -;
PIR; T30333; T30333.
SQ SEQUENCE 129 AA; 14339 MW; 19ABBB84968ECE9 CRC64;
Query Match 38.3%; Score 273.5; DB 2; Length 129;
Best Local Similarity 43.1%; Pred. No. 2e-21;
Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 3;
QY 1 MTSSQQRVERFLOYSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSLSLLHCRRIE 60
Db 1 MKTSQDFARFINSIGLQGLTSLQNGVCALYDQNEAAIIEPHESEWVFHCRIGR 60
QY 61 ADPQTSITLYSMLLQNFEMAMRGCGWLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFI 120
Db 61 C-PERAPDLL-RLLSLNFDFVARLHGCNFAVDQ-GDVRLCAQRELASLDEPAFCDVTRGFI 117
QY 121 EHAAEVREYI 130
Db 118 SQAREARAPL 127
RESULT 5
Q9JP37 PRELIMINARY; PRT; 131 AA.
ID Q9JP37
AC Q9JP37;

DR GO: 0004350; F-glutamate-5-semialdehyde dehydrogenase acti. . . ; IEA.
 DR GO: 0005561; P-proline biosynthesis; IEA.
 DR InterPro; IPR000965; Gglut_pp_reduct.
 DR PROSITE; PS01223; PROA; 1.
 KW Complete proteome.
 SQ SEQUENCE 423 AA; 45417 MW; E2B9930ABE513A11 CRC64;

Query Match 11.4%; Score 81.5; DB 16; Length 423;
 Best Local Similarity 20.5%; Pred. No. 2.9; Indels 63; Gaps 8;
 Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;
 QY 12 LQYFSAAGCTPI--HLKDGVCALYNEQDEAAVLEVPQHSQ-----VSQADLLKANNVAFNAKTYRYGIC 260
 DB 213 IERISRDARVPVVKHL-DGICHYI-----VSQADLLKANNVAFNAKTYRYGIC 260
 QY 51 ----SILLHCRITADPQTSITLYSMLLQNFEMAMRG-----WLA- 89
 DB 261 GAMETLLVQQAER-----FLPEMARFRVEKRGVLCGRTQAIISAKPATEADWHITE 314
 QY 90 -LDELHNVLC--FQSLHLEASFSDIVSGFIEHAAEVREYIAQLDESS 137
 DB 315 YLDAILLSIRVDGNGINQAIHINHYGSHHTDSIIEHQGEARQPMFVDSAS 365

RESULT 8

Q92R07 PRELIMINARY; PRT; 186 AA.
 AC Q92R07;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein R00802.
 GN R00802 OR SMC00902.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puchler A., Purnelle B., Rampeger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 Sinorhizobium meliloti strain 1021";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
 DR EMBL: AL591785; CAC45374.1; --
 DR GO: 0003677; P:DNA binding; IEA.
 DR GO: 0008170; F:N-methyltransferase activity; IEA.
 DR GO: 0006306; P:DNA methylation; IEA.
 DR InterPro; IPR004398; Cons.hypoth95.
 DR InterPro; IPR002052; N6 MCase.
 DR Pfam; PF03602; Cons.hypoth95; 1.
 DR TIGRFAMs; TIGR00095; TIGR00095; 1.
 DR PROSITE; PS00092; N6_MCASE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 186 AA; 20282 MW; 19B53236451C6EE6 CRC64;

Query Match 11.3%; Score 81; DB 16; Length 186;
 Best Local Similarity 25.0%; Pred. No. 1.3;
 Matches 34; Conservative 19; Mismatches 41; Indels 42; Gaps 5;
 QY 12 LQYFSAAGCTPIHLKDGVCALYNEQDEAAVLEVPQHSQSLHCR----- 57
 DB 61 LEALSRCGRQVLFVEQGV-----EGGLIRI--NIEALGQGRKIFRRDATDLGFV 110
 QY 58 -----IIEADPQTSITLYSMLLQNFEMAMRGWLA-----LDELHNVLCFQOQL 104
 DB 111 GTMEPFHLVADPPYKGLGERALS-----AARGGMLVPGALALIEERADVRPQFSEF 165

QY 105 EHLDEASFSDIVSGFI 120
 DB 166 ESDERAFGDTLMHFL 181
 RESULT 9
 Q8ZW74 PRELIMINARY; PRT; 723 AA.
 ID Q8ZW74;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Carbon monoxide dehydrogenase large subunit, conjectural.
 GN PAE1935.
 OS Pyrobaculum aerophilum.
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
 OC Thermoproteaceae; Pyrobaculum.
 OX NCBI_TaxID=13773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
 RX MEDLINE=21664397; PubMed=11792869;
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
 RA Miller J.H.;
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
 aerophilum";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989 (2002).
 DR EMBL: AF009848; AAL63828.1; --
 DR GO: 0016491; F:oxidoreductase activity; IEA.
 DR GO: 0006118; P:electron transport; IEA.
 DR InterPro; IPR008274; Aldxan_dh_hamm.
 DR InterPro; IPR000674; Aldxan_dh_hamm.
 DR Pfam; PF01315; Ald_Xan_dh_C2; 1.
 DR Pfam; PF02738; Ald_Xan_dh_C2; 1.
 KW Complete proteome.
 SQ SEQUENCE 723 AA; 78914 MW; FCA65FADDE9FB235 CRC64;
 Query Match 11.3%; Score 81; DB 17; Length 723;
 Best Local Similarity 24.6%; Pred. No. 6.1;
 Matches 34; Conservative 27; Mismatches 53; Indels 24; Gaps 6;
 QY 15 FSAAGCTPIHLKDGVCALYNEQDEAAVLEVPQHSQSLHCRITIEADPQTSITLYS 71
 DB 537 FSCGKGSAT-VKDVVRVAYGEAAQTVEAIYHADSTFFGVHIAVLELDPEIGF-VKP 594
 QY 72 MLLQLNFEMAMRGWLADELHNVLCFQOQSLHLEASFSDIV---SG----- 118
 DB 595 MLKSYDDVGVVNPPLASQITG-----GALQGIQAALYEEVYDESGNLTSLNLA 648
 QY 119 FIEHAAEVREYIAQLDES 136
 DB 649 YVPTAAEAPKVEYVFAES 666
 RESULT 10
 Q40879 PRELIMINARY; PRT; 368 AA.
 ID Q40879;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Myristyl-ACP desaturase.
 OS Pelargonium hortorum (Common geranium).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC Geraniales; Geraniaceae; Pelargonium.
 OX NCBI_TaxID=4031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=line: 88-51-10; TISSUE=Trichome gland;
 RX MEDLINE=96323300; PubMed=8710947;
 RA Schultz D.J., Cahoon E.B., Shanklin J., Craig R., Cox-Foster D.L.,
 RA Mumma R.O., Medford J.I.;


```

RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL: BX248342; CAD97368.1; -.
KW Transferase; Acyltransferase; Complete proteome.
SQ SEQUENCE 789 AA; 88340 MW; 85ACA549B9B1CFF1 CRC64;

Query Match 11.0%; Score 79; DB 16; Length 789;
Best Local Similarity 24.3%; Pred. No. 11;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

QY 23 IHKQGVCAALYNEQDEAAVLEVPQHSDDLHLCRIIEADPQTSITLYSMLLQINFEMAA 82
DB 445 VSMRQVLGAPGHELTQDPAKRLALQKMSPEVAVIRILQATPVATGUVSALL-----LT 498
QY 83 MRCQWALDELHNVRLCFOQSLHLD-----EASFSDIVSGF----- 119
DB 499 TRGTALTDQGHHT--LQSDLDYLRKQSPVSTSLRLRSREGVRAAALSNGHPVTR 555
QY 120 IEHAAEVREYIAQLDESSAA 139
DB 556 VDSGREFVWIAPDDEHAAA 575

RESULT 14
Q9PE32 PRELIMINARY; PRT; 830 AA.
AC Q9PE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Ribonucleoside-diphosphate reductase large chain (EC 1.17.4.1).
GN XF1196.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lenos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.O.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

```

```

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC LARGE CHAIN FAMILY.
DR EMBL: AE003954; AAF84006.1; -.
DR PIR: C82710; C82710.
DR GO: GO:0005971; C:ribonucleoside-diphosphate reductase complex; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0004748; F:ribonucleoside-diphosphate reductase activity; IEA.
DR GO: GO:0006260; P:DNA replication; IEA.
DR InterPro: IPR000788; Ribonucleo_red.
DR InterPro: IPR008926; Ribonucleo_red_N.
DR Pfam: PF00317; ribonuc_red_lg; 1.
DR Pfam: PF02867; ribonuc_red_lgc; 1.
DR PRINTS: PR01183; RIBORDTASEM1.
DR PROSITE: PS00089; RIBORED_LARGE; 1.
KW DNA replication; Oxidoreductase; Complete proteome.
SQ SEQUENCE 830 AA; 94229 MW; 12207407246E28BC CRC64;

Query Match 11.0%; Score 78.5; DB 16; Length 830;
Best Local Similarity 25.6%; Pred. No. 13;
Matches 33; Conservative 21; Mismatches 46; Indels 29; Gaps 6;

QY 17 AGCKTPHLKDGVCALYNEQDEAAVLEVPQHSDDLHLCRII-EADPQTSITLYSMLLQ 75
DB 22 AGSVPPVALS-----VFHPDEPD-----EVPIPRQTMAMHTRAVEANVSTWITKEAGNR 73
QY 76 LNFEMAAMRCGLALDELHNVRLCFOQSLHLDSEASFDIVSGFIEHAAEV----- 126
DB 74 MPFDCARLE---CAIDTIH-----QEFQLDVAEYKRAVGFVEVERKDSVNADDIVDLL 123
QY 127 -REYIAQLD 134
DB 124 IREAEARVD 132

RESULT 15
Q8RP06 PRELIMINARY; PRT; 155 AA.
AC Q8RP06;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Putative chaperone ChpPmak.
OS Pseudomonas syringae pv. maculicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=59511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ES4326;
MEDLINE=21862332; PubMed=11872842;
RA Guttman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA Greenberg J.F.;
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant
RT Pathogen Pseudomonas syringae.";
RL Science 295:1722-1726(2002).
DR EMBL: AF458049; AAL84250.1; -.
SQ SEQUENCE 155 AA; 17124 MW; 33EF90778D00E809 CRC64;

Query Match 10.9%; Score 78; DB 2; Length 155;
Best Local Similarity 26.3%; Pred. No. 2.1;
Matches 31; Conservative 21; Mismatches 48; Indels 18; Gaps 5;

QY 17 AGCKTPHLKDGVCALYNEQDEAAVLEVPQHSDDLHLCRII-EADPQTSITLYSMLLQ 76
DB 24 AGCVD-----QGHSA-FSDREGNLFQAVPDPAFSSLLHYLTLELQAVDGRVLASALAI 78

```


This Page Blank (uspto)